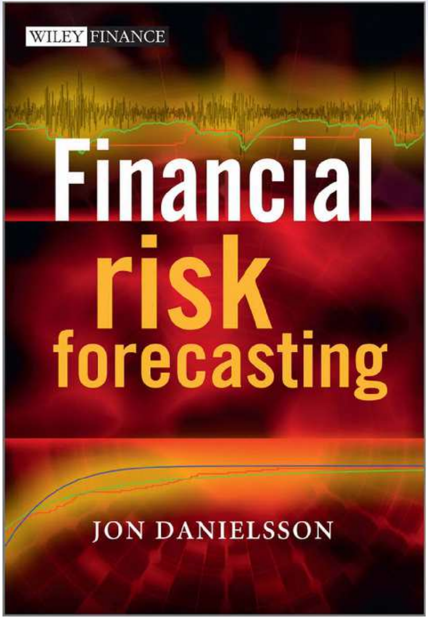
Financial Risk Forecasting
Chapter 1
Financial Markets, Prices and Risk

Jon Danielsson ©2025
London School of Economics

To accompany
Financial Risk Forecasting
FinancialRiskForecasting.com

Published by Wiley 2011

Version 10.0, August 2025


WILEY FINANCE
Financial
risk
forecasting
JON DANIELSSON

Introduction To Chapter

Financial Markets, Prices and Risk

- Based on Chapter 1 in *Financial Risk Forecasting* with updated data
- Statistical techniques for analysing prices and returns
- Stock market indices, for example the S&P-500
- Prices, returns and volatilities – Three stylised facts:
  1. Volatility clusters
  2. Fat tails
  3. Non-linear dependence
- See Appendix A of *Financial Risk Forecasting* for more detailed discussion on the statistical methods
- Introduction to simulations
- Case: Covid-19

Notation new to this Chapter

T	Sample size
t	A particular observation period (eg a day)
p_t	Price at time t
r_t	Simple return
y_t	Continuously compounded return
σ	Unconditional volatility
σ_t	Conditional volatility
μ	Mean
K	Number of assets
ρ	Probability
q	Quantile
w	$K \times 1$ vector of portfolio weights
ν	Degrees of freedom of the Student-t
d	Dividends

Risk Is Latent

- A farmer feeds his turkeys every day at 7am
- A scientist turkey discovers:
  - "Food arrives every morning at 7am"
  - And tells this to the other turkeys
- On Thanksgiving morning the farmer comes but does not bring food
- Instead, he slaughters all the turkeys

Lesson: If you only model and forecast risk by looking at past prices you will eventually run into a situation where a huge loss will completely surprise you.

Technically, risk is a latent variable that cannot be measured directly and can only be inferred from observed data.

Prices, Returns and Indices

Stock Indices

- A *stock market index* shows how a representative portfolio of stock prices changes over time
- A *price-weighted* index weighs stocks based on their prices
  - A stock trading at $100 makes up 10 times more of total than a stock trading at $10
- A *value-weighted* index weighs stocks according to the total market value of their outstanding shares
  - Impact of change in stock price proportional to overall market value

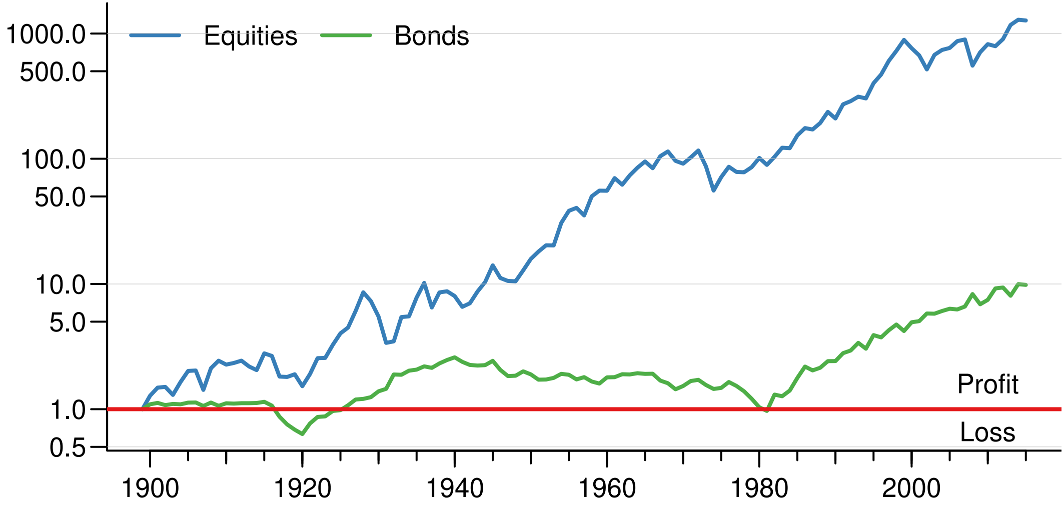
Examples

- The most widely used index is the Standard & Poor's 500 (*S&P-500*) – the 500 largest traded companies in the US
- Examples of value-weighted indices
  - S&P-500, FTSE 100 (UK), TOPIX (Japan)
- Examples of price-weighted indices
  - Dow Jones Industrial Average (US), Nikkei 225 (Japan)

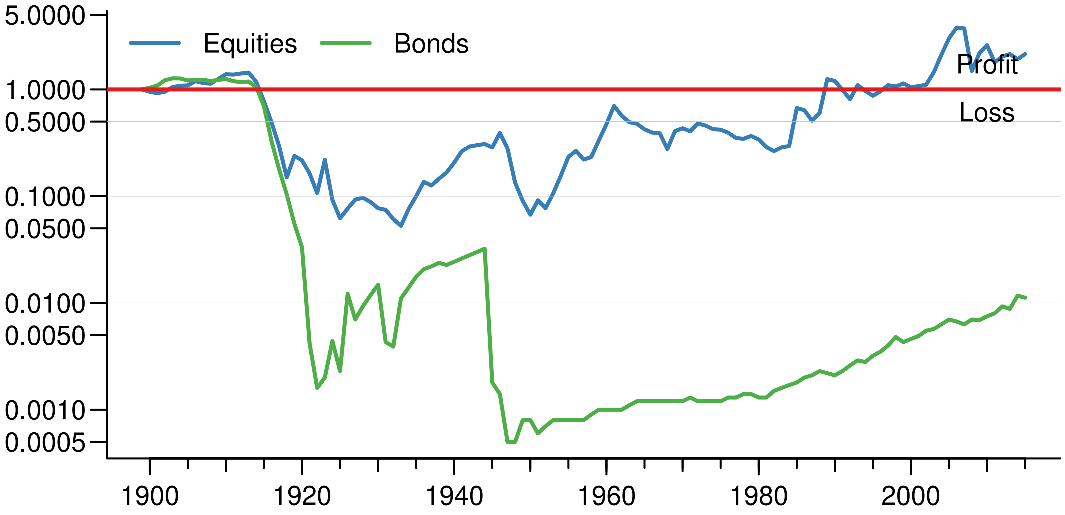
Total Return Indices

- A stock market index shows how prices evolve
- However, investors would earn more because some of the stocks pay dividends – They can have splits and buybacks
- A *total return index* incorporates all of these
- If you only adjust for number of shares, we get what is sometimes called *adjusted returns*
- However, inflation is still excluded
- And that can make price comparisons across long time periods invalid
- Unless we adjust for inflation
- Sometimes, when we use the phrase "total returns", for adjusted returns inflation and dividends are included
- Like in the next plots

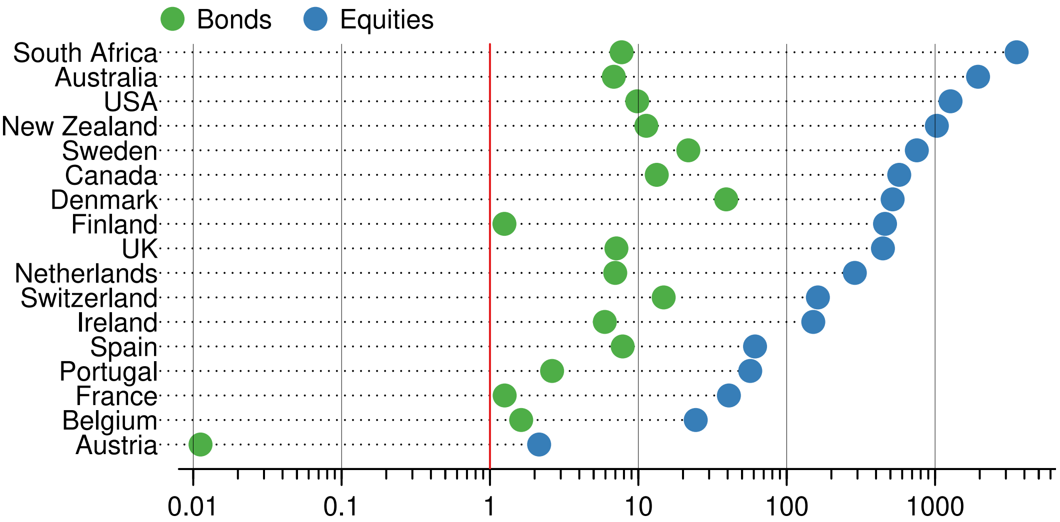
Total Returns 1900 to 2016: USA



Total Returns 1900 to 2016: Austria



Total Returns 1900 to 2016



Bonds and Equities

- The very long-run returns on equities are much higher than those on bonds
- But bonds can have a very good short- and medium-term performance
- Especially if inflation and hence interest rates are falling
- That was the case in many countries from the early 1980s until the early 2020s

Prices and Returns

- Denote prices on day t by p_t
- Usually we are more interested in the *return* we make on an investment

Definition return The relative change in the price of a financial asset over a given time interval, often expressed as a percentage

- There are two types of returns
  1. *Simple*: (r)
  2. *Compound* or *log*: (y)

Simple Returns

Definition A simple return is the percentage change in prices.

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1$$

- Including dividends d

$$r_t = \frac{p_t - p_{t-1} + d}{p_{t-1}}$$

$$p_{t+1} = p_t(r_t + 1)$$

Continuously Compounded Returns

Definition The logarithm of gross return.

$$y_t = \log(1 + r_t) = \log\left(\frac{p_t}{p_{t-1}}\right) = \log(p_t) - \log(p_{t-1})$$

$$p_{t+1} = p_t e^{y_t}$$

Simple and Continuous

- The difference between r_t and y_t is not large for daily returns
- As the time between observations goes to zero, so does the difference between the two measures

$$\lim_{\Delta t \to 0} y_t = r_t$$

$$\log(1000) - \log(990) = 0.01005 \approx \frac{1000}{990} - 1 = 0.0101$$

$$\log(1000) - \log(800) = 0.223 \neq \frac{1000}{800} - 1 = 0.25$$

Symmetry

- Continuous returns are *symmetric*

$$\log\left(\frac{1000}{200}\right) = -\log\left(\frac{200}{1000}\right)$$

- Simple are not

$$\frac{1000}{200} - 1 \neq -\left(\frac{200}{1000} - 1\right)$$

Issues for Portfolios

- $r_{t,\text{portfolio}}$ return on a portfolio
- Weighted sum of returns of K individual assets:

$$r_{t,\text{portfolio}} = \sum_{k=1}^{K} w_k r_{t,k} = w' r_t$$

- While

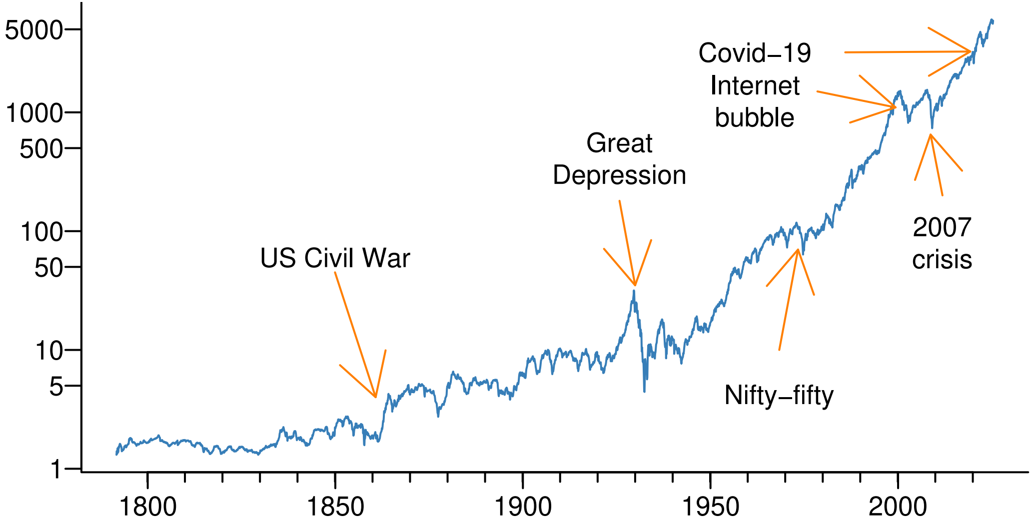
$$y_{t,\text{portfolio}} = \log\left(\frac{p_{t,\text{portfolio}}}{p_{t-1,\text{portfolio}}}\right) \neq \sum_{k=1}^{K} w_k \log\left(\frac{p_{t,k}}{p_{t-1,k}}\right)$$

- Because the log of a sum does not equal the sum of logs

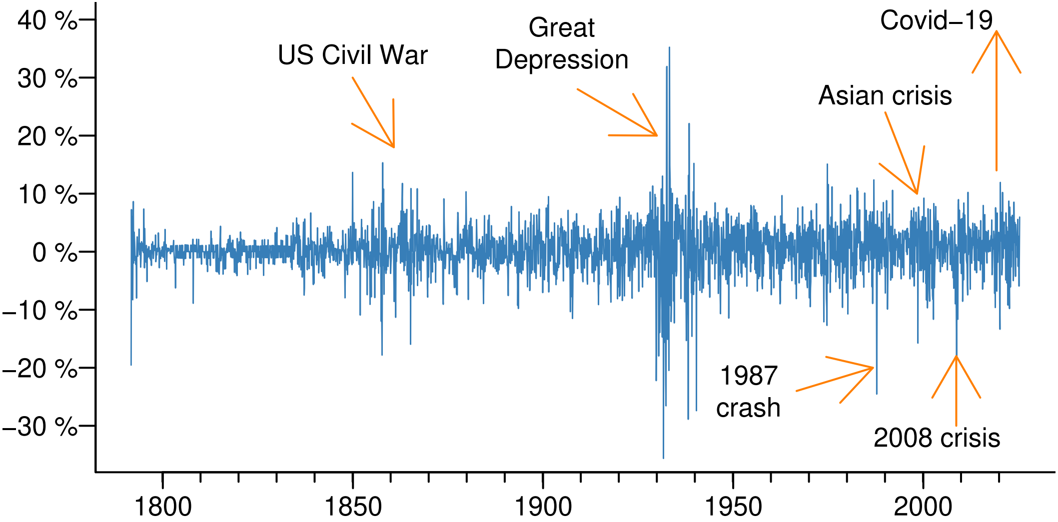
Comparison

- Simple returns are
  - Used for accounting purposes
  - Investors are usually concerned with simple returns
- Continuously compounded returns have some advantages
  - Mathematics is easier (for example, how returns aggregate over many periods, used in Chapter 4)
  - Used in derivatives pricing, for example, the Black-Scholes model

S&P-500 Index Sep 1791 - May 2025



S&P-500 Monthly Returns Sep 1791 - May 2025



S&P-500 Statistics

1929 to April 2024, daily returns

Mean	0.0307%
Standard error	1.366%
Min	−22.90%
Max	15.37%
Skewness	−0.528
Kurtosis	21.55

- Note how small mean is compared to the standard error (volatility)
- But mean grows at rate T and the volatility at $\sqrt{T}$ which becomes important later

Three Stylised Facts: Present in Most Financial Returns

a. Volatility clusters
b. Fat tails
c. Non-linear dependence

Review of Relevant Statistics

Probability Density Function (PDF)

- A PDF, $f(x)$, describes likelihood of outcomes for random variables
- Area under the curve is probability
- Probability between two values:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$$

- Total area under a *standard* PDF is 1:

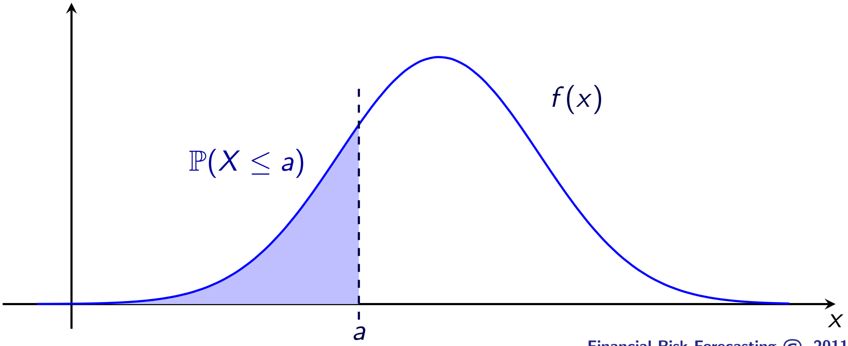
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

- The standard normal has variance 1, but other distributions may not

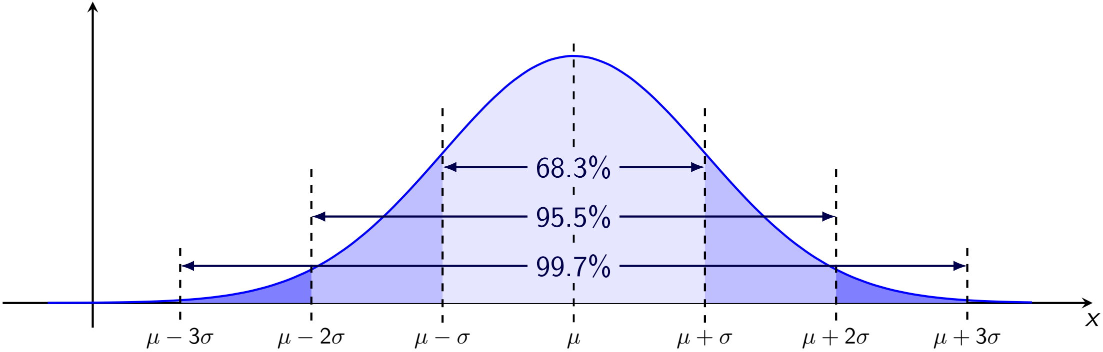
Normal (Gaussian) Probability Density Function (PDF)

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

$$f(x;\mu,\sigma) = \phi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$



Area Under the Normal



Cumulative Distribution Function (CDF)

- A CDF, $F(q)$, gives the probability that $x \leq q$
- It is defined as:

$$F(q) = \int_{-\infty}^{q} f(t)\, dt$$

- Rises from 0 to 1 as q increases
- Always non-decreasing and right-continuous
- Useful for computing quantiles and probabilities
- For continuous distributions:

$$f(q) = \frac{d}{dx} F(q)$$

Quantiles

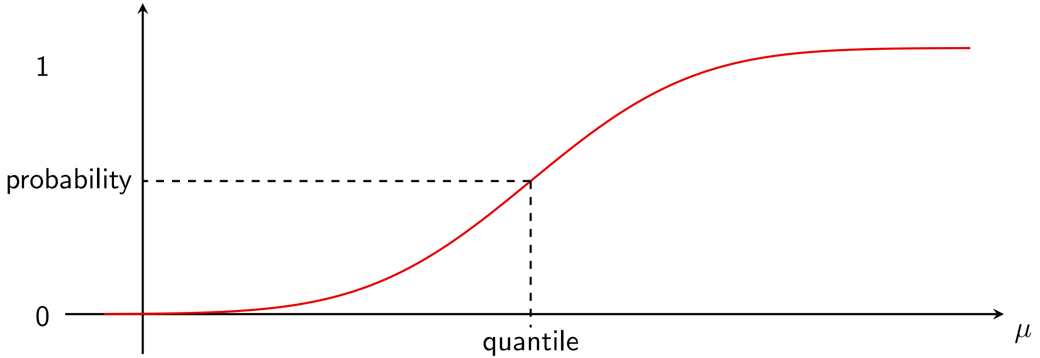
- ρ is probability, $0 \leq \rho \leq 1$
- Can express in percentages, so $0\% \leq \rho\% \leq 100\%$ or $0\% \leq \rho \leq 100\%$ or $0 \leq \rho \leq 100$
- A quantile is a point below which a given proportion of data falls
- The ρ-quantile, q_ρ, satisfies: $\mathbb{P}(x \leq q_\rho) = \rho$
- Common examples:
  - Median: $\rho = 0.5$
  - Lower quartile: $\rho = 0.25$
  - 1% quantile: $\rho = 0.01$, used in value-at-risk in Chapter 4
- Useful for describing the distribution of returns, especially tails

Quantiles and Probabilities

- The quantile function is the inverse of the cumulative distribution function (CDF)
- Given ρ, the ρ-quantile q_ρ satisfies: $F(q_\rho) = \rho$
- Given a value x, the probability is: $F(x) = \mathbb{P}(x \leq q)$
- So:
  - Start with ρ, compute $q_\rho = F^{-1}(\rho)$
  - Start with x, compute $\rho = F(x)$
- Quantiles turn probabilities into thresholds and vice versa

Cumulative Normal Cumulative Distribution Function (CDF)

$$\Phi(x) = \int_{-\infty}^{x} \phi(x)\, dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}\right) dt$$



Moments, Means, Variances and Standard Deviation

- Moments are the expected value of a random variable to some power
- So the m^{th} moment is:

$$\mathrm{E}[X^m] = \int_{-\infty}^{\infty} x^m f(x) dx$$

- The mean is the first moment

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- The variance is a function of the first and the second moment

$$\mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] + \mu^2 - 2\mu\,\mathrm{E}[X]$$

- Standard deviation is the square root of variance

Covariance and Correlation

- Covariance measures joint variability of two variables

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) \, dx \, dy$$

- Positive covariance: variables move together
- Negative covariance: one rises, the other falls
- Correlation is *standardised* covariance

$$\frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Correlation ranges from -1 to $+1$
- Correlation $= 0$ implies no **linear** relationship

Skewness

- Skewness measures asymmetry of a distribution
- Defined as the third standardised moment:

$$\text{Skewness} = \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - \mu)^3 f(x)\, dx$$

- Skewness > 0: longer right tail (positive skew)
- Skewness < 0: longer left tail (negative skew)
- Skewness $= 0$: symmetric distribution

Kurtosis

- Kurtosis measures tail weight and peak sharpness (figure a bit later)
- Defined as the fourth standardised moment:

$$\text{Kurtosis} = \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - \mu)^4 f(x) \, dx$$

- Normal distribution has kurtosis $= 3$
- Excess kurtosis $=$ Kurtosis -3
- High kurtosis: fat tails, more outliers
- Low kurtosis: light tails, fewer outliers
- What about normal (gaussian) tails?

Fat Tails

- A distribution is fat-tailed if it has more extreme outcomes than a normal with the same mean and variance
- Discuss later what that implies for the mean-variance (MV) model

Sample Estimators

- Mean:

$$\hat{\mu} = \frac{1}{n}\sum x_i$$

- Variance:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum (x_i - \hat{\mu})^2$$

- Standard deviation:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

- Covariance

$$\widehat{\text{Cov}}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- Skewness:

$$\frac{1}{n}\sum\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^3$$

- Kurtosis:

$$\frac{1}{n}\sum\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^4$$

- Correlation:

$$\frac{\widehat{\text{Cov}}(x,y)}{\hat{\sigma}_x \hat{\sigma}_y}$$

What Is a Financial Time Series?

- A time series is data observed over time at regular or irregular intervals
- In finance: daily prices, returns, interest rates, volatility
- Time order matters: yesterday affects today
- Typical patterns:
  - Trend in prices
  - Volatility clustering in returns

AR and MA Models

Autoregressive (AR) model:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \varepsilon_t$$

- Captures momentum or memory in the series itself

Moving average (MA) model:

$$x_t = b_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

- Captures persistence in shocks (e.g., after a surprise)

AR looks at past values. MA looks at past surprises

Autocorrelations

- Correlations measure how two variables (x, y) move together

$$\text{Corr}(x, y) = \frac{\sum_{t=1}^{T}(x_t - \mu_x)(y_t - \mu_y)}{(T-1)\sigma_x\sigma_y}$$

- Autocorrelations measure how a single variable is correlated with itself at different lags
  - *1* lag

$$\hat{\beta}_1 = \text{Corr}(x_t, x_{t-1})$$

  - i lags

$$\hat{\beta}_i = \text{Corr}(x_t, x_{t-i})$$

R

```
acf(y,20)
```

The Ljung-Box (LB) Test for Autocorrelations

- Joint significance of autocorrelation $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N)$ can be tested by using the Ljung-Box (LB) test
- It is χ^2 distributed because a normal squared is distributed χ^2 and we assume the data is normally distributed
- And the degrees of freedom arises from testing multiple, N, lags at the same time

$$J_N = T(T+2)\sum_{i=1}^{N}\frac{\hat{\beta}_i^2}{T-N} \sim \chi^2_{(N)}$$

R

```
Box.test(y, lag = 20, type = c("Ljung-Box"))
```

Stationarity

- A time series is weakly stationary if:
  - Mean is constant over time
  - Autocovariance depends only on the lag, not time
  - Unconditional variance exists and is constant
- Stationarity is required for most models like AR and MA
- Prices are not stationary
- Returns are often weakly stationary, but may have time-varying volatility.
- GARCH models assume a stationary mean, but allow the conditional variance to change over time

Stationarity means we can learn from the past — without it, models chase moving targets

p-values vs. Critical Values in Hypothesis Testing

- Critical value approach
  - Set significance level α (e.g., 0.05)
  - Compute test statistic
  - Reject H_0 if statistic exceeds critical value
- p-value approach
  - p-value $=$ probability of observing result at least as extreme as test statistic
  - Reject H_0 if p-value $< \alpha$
- Both methods give the same decision

Normal Squared and the Chi-Squared Distribution

- If $x \sim \mathcal{N}(0,1)$, then $x^2 \sim \chi^2_1$, where the subscript 1 indicates one degree of freedom
- More generally, the sum of k independent squared standard normals:

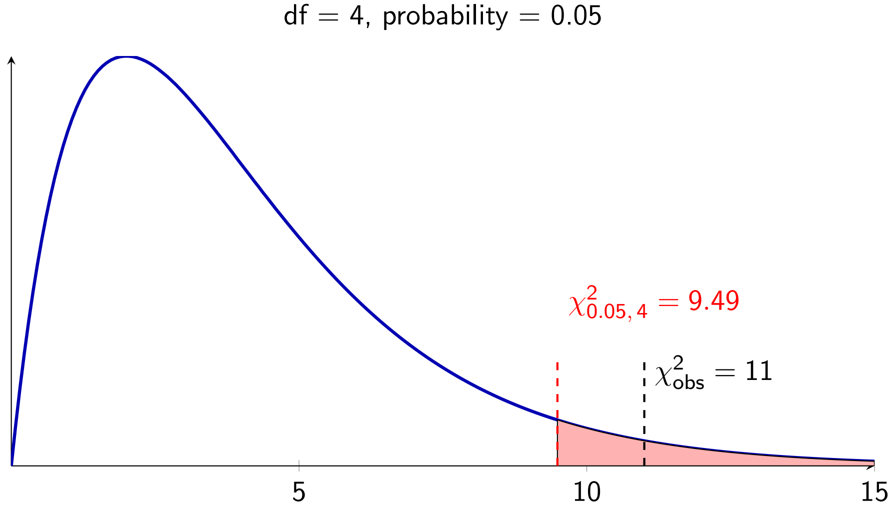
$$\sum_{i=1}^{k} x_i^2 \sim \chi^2_k$$

- The chi-squared distribution is skewed and only takes non-negative values
- Used extensively in statistical testing:
- Critical values from the χ^2 distribution determine test rejection regions

Chi-squared (χ^2) Hypothesis Testing

- Null Hypothesis (H_0): Observed data fits expected distribution
- Alternative Hypothesis (H_1): Observed data does *not* fit
- Compare observed test statistic to critical value from χ^2 distribution table
- If $\chi^2_{\text{observed}} > \chi^2_{\text{critical}}$, reject H_0
- Or use p-value. If p-value $<$ threshold (e.g. 5%), reject H_0

Chi-squared Hypothesis Testing



Discrete vs. Continuous Random Variables

- Continuous random variables
  - Take on values in a *continuous range* (uncountable).
  - Height of a person, time until failure of a device.
  - Price of a stock
- Discrete random variables (Chapter 4 and 8)
  - Take on a countable number of distinct values
  - Number of heads in 10 coin tosses, number of students in a class
  - A bond that either pays out or defaults (see Chapter 4)
  - Value-at-Risk violation (see Chapter 8)

Volatility

Volatility: The Standard Deviation/Error of Returns

- Two concepts of volatility:
  - *Unconditional volatility* is volatility over an entire time period (σ)
  - *Conditional volatility* is volatility in a given time period, conditional on what happened before (σ_t)
- σ vs σ_t
- The subscript t tells us it is the volatility of a particular time period, in this course usually a day
- Clear evidence of cyclical patterns in volatility over time, both in the short run and the long run
- Volatility is risk if and only if returns are normally distributed

Calculations

- Daily volatility (mean is μ)

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2}$$

- Annualised

$$\sqrt{250}\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2}$$

- Why 250 and not 365? Because the 250 is a typical number of days the market is open per year (trading days)

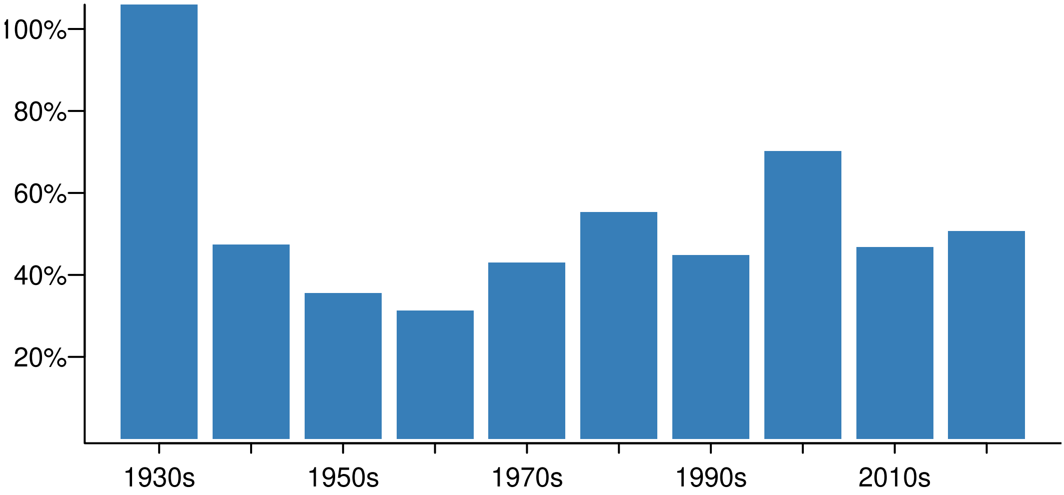
R

```
sd(y)
sqrt(250)*sd(y)
```

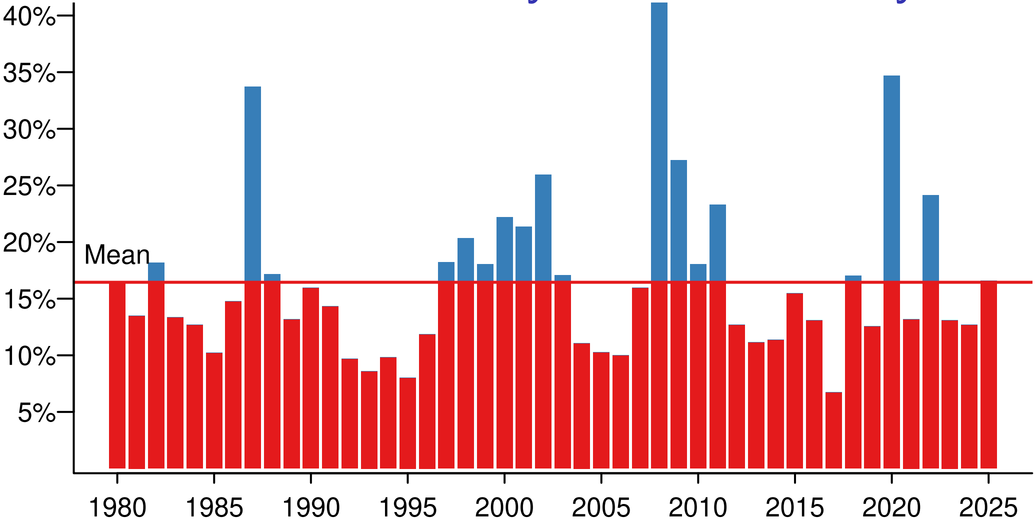
Volatility Clusters

- Suppose we use the annualised volatility equation and calculate volatility over a decade, year and month, using daily returns (a method called realised volatility)
- Then we see that volatility comes in many cycles
- Both long-run and short-run
- We call these *volatility clusters*

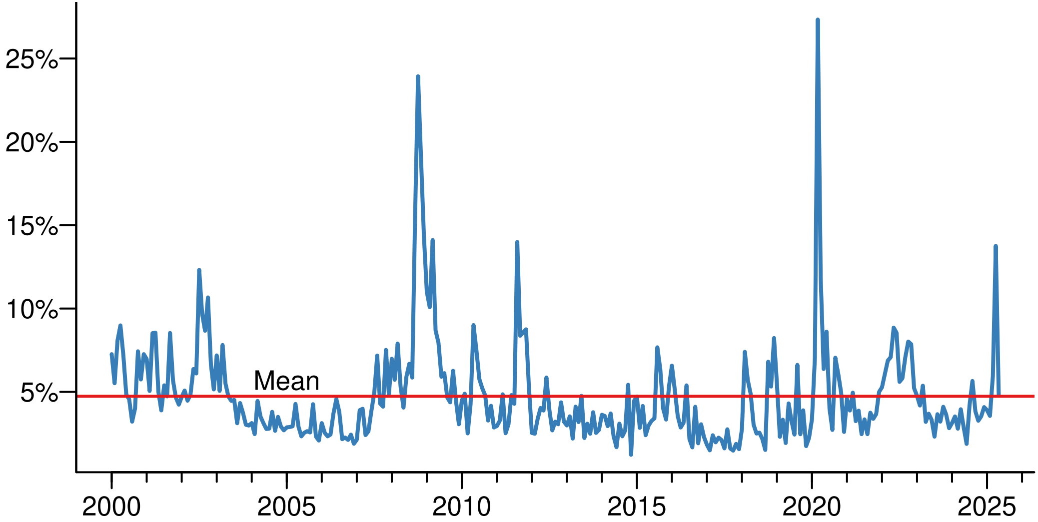
Clusters in S&P-500 Volatility: Decade



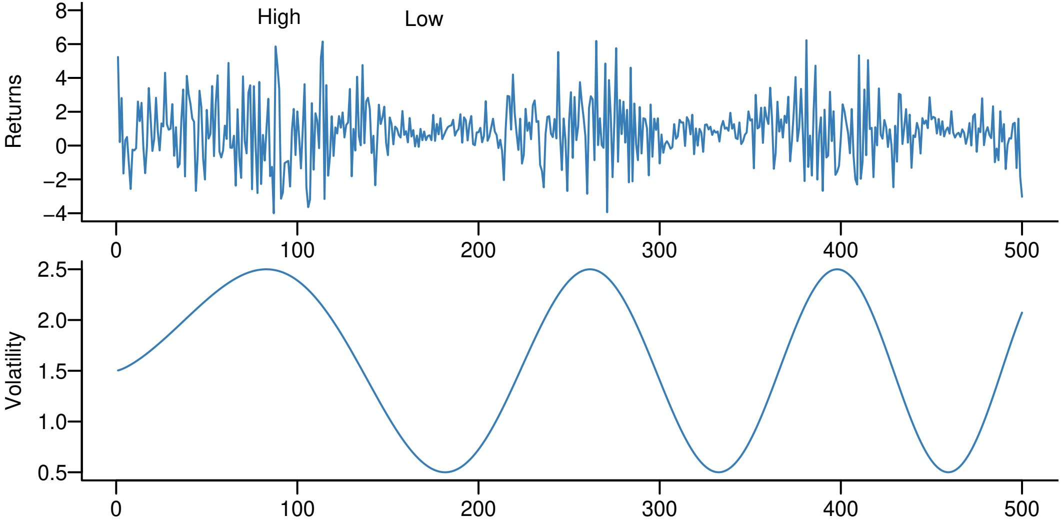
Clusters in S&P-500 Volatility: Annual 1980 - May 2025



Clusters in S&P-500 Volatility: Monthly 2000 - May 2025



Simulated Volatility Clusters



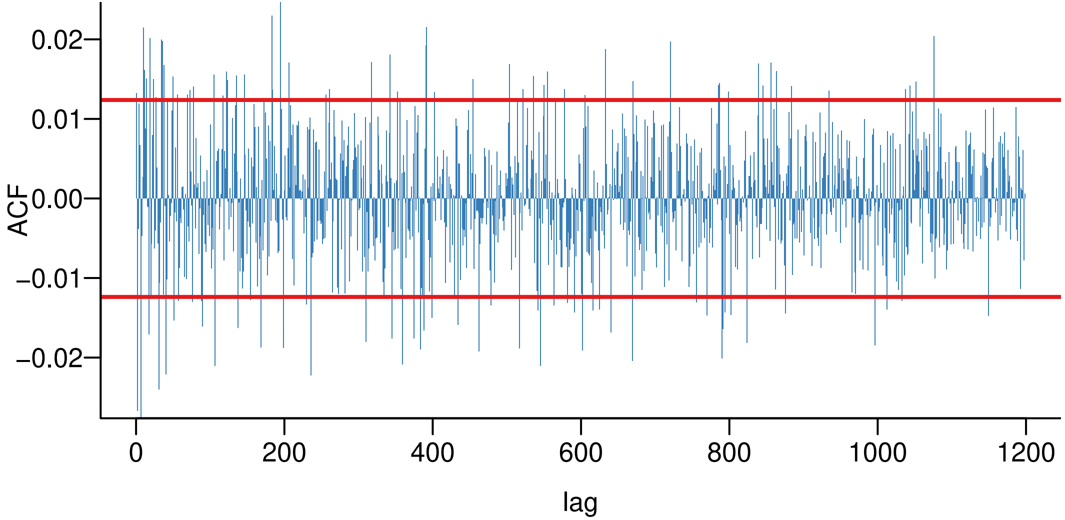
Volatility Clusters

- Volatility changes over time in a way that is partially predictable
- *Volatility clusters*
- Engle (1982) suggested a way to model this phenomenon
  - His autoregressive conditional heteroskedasticity (ARCH) model is discussed in Chapter 2

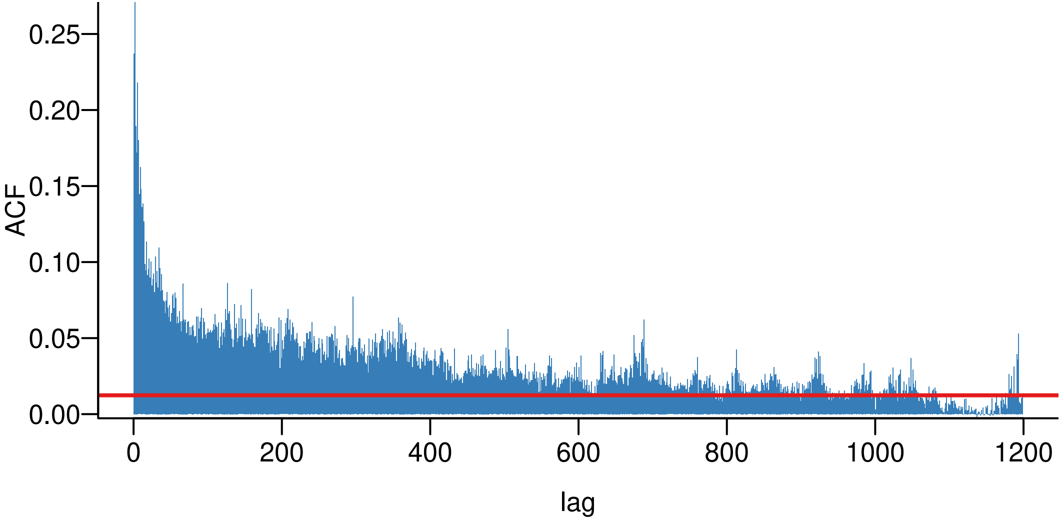
Autocorrelations (cont.)

- If autocorrelations are statistically significant, there is evidence for predictability
- The coefficients of an autocorrelation function (ACF) give the correlation between observations and lags
- We will test both returns (y), predictability in mean (price forecasting or alpha)
- And squared returns, which capture predictability in volatility

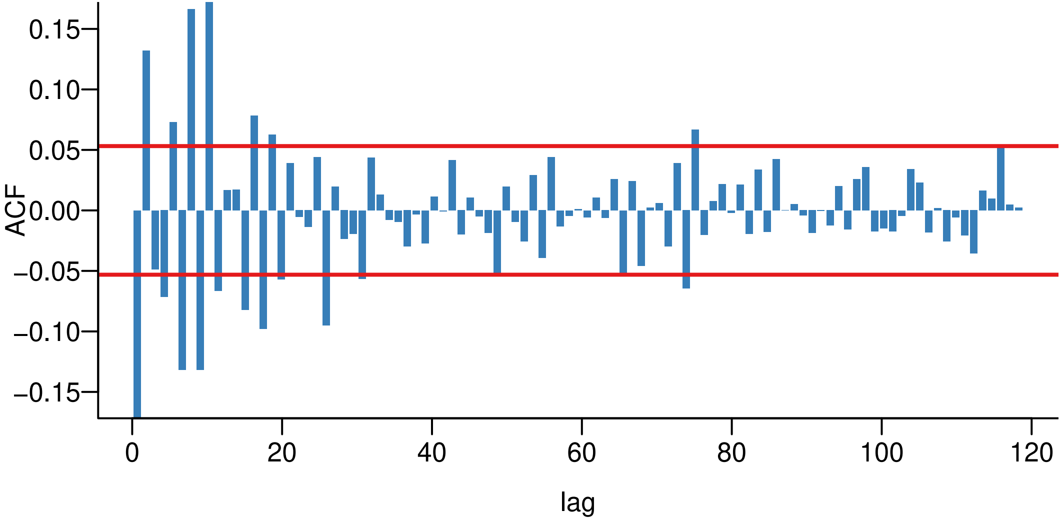
S&P-500 ACF of Daily Returns. 1929 - May 2025



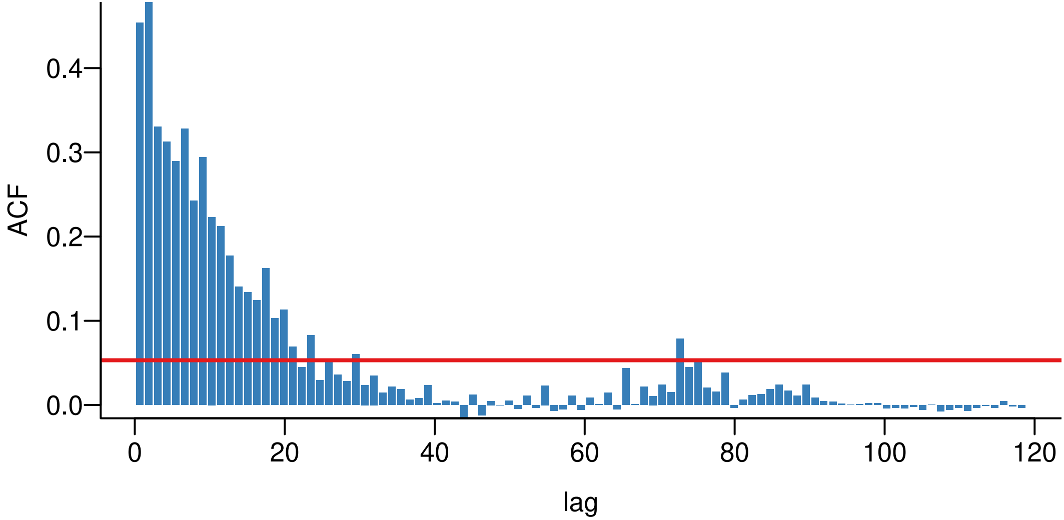
S&P-500 ACF of Squared Daily Returns. 1929 - May 2025



S&P-500 ACF of Daily Returns. 1929 - May 2025



S&P-500 ACF of Squared Daily Returns. 1929 - May 2025



LB Tests for S&P-500

Daily returns

N	LB statistic, 21 lags	p-value
22,752	95.9	1.527×10^{-11}
2,500	185.2	$< 2.2 \times 10^{-16}$
100	18.7	0.606

Daily returns squared

T	LB statistic, 21 lags	p-value
22,752	12,633.0	$< 2.2 \times 10^{-16}$
2,500	4,702.1	$< 2.2 \times 10^{-16}$
100	46.0	0.00129

Market Efficiency

- In an efficient market, prices reflect all available information
- Any predictable profit opportunity above costs is quickly competed away
- Costs include trading fees, taxes, bid-ask spreads and opportunity costs (like investing in a risk free asset)
- Three forms of market efficiency:
  - Weak form: prices reflect past market data
  - Semi-strong form: prices reflect all public information
  - Strong form: prices reflect all public and private information
- Persistent outperformance may suggest either hidden risk or market inefficiency

What Do the Results Say About Market Efficiency?

- Weak form efficiency suggests past price movements, volume and earnings data do not affect a stock's price sufficiently strongly to allow one to systematically make money predicting future prices
- The ACF is (almost) insignificant for the mean, when taking into account the risk free rate, inflation and trading costs
- The ACF is very significant for returns squared, suggesting volatility is highly predictable
- That does not mean market efficiency is violated
- As the cost of carry, cost of holding a security over a period of time, is very high for volatility products
- Neither result suggests *one cannot* make money forecasting prices or volatilities
- Nor do they suggest *one can*

Fat Tails

Definition

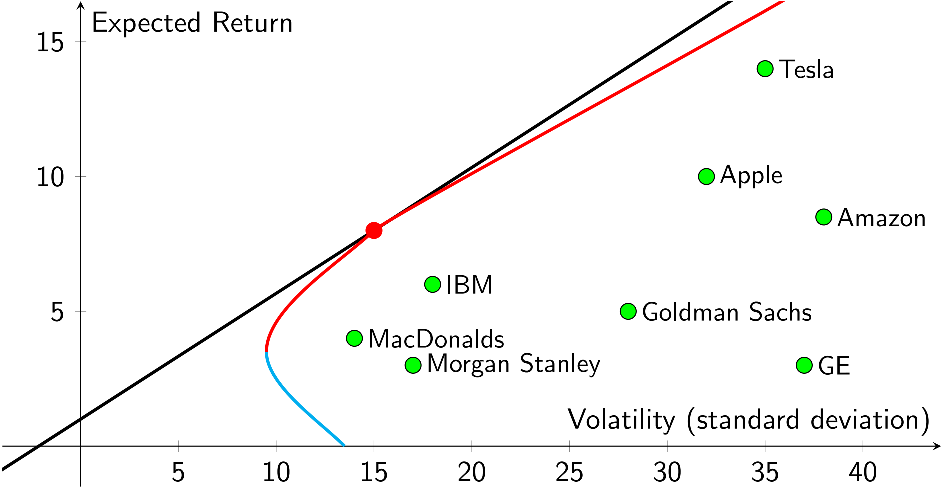
Fat tails: A random variable is said to have fat tails if it exhibits more extreme outcomes than a normally distributed random variable with the same mean and variance.

- The mean–variance model assumes normality (next 2 slides)

Mean-Variance Model

- Developed by Harry Markowitz in 1952
- Helps investors build portfolios by balancing
  - expected return – mean
  - risk – variance
- Assumes investors prefer more return and less risk
- Leads to the efficient frontier — a set of optimal portfolios that offer the best possible return for a given level of risk
- It assumes normality because when only mean and variance matter, the normal distribution is the only one fully characterised by these two parameters

Efficient Frontier (made up numbers)



Fat Tails

- The tails are the extreme left and right parts of a distribution
- If the tails are fat, there is a higher probability of extreme outcomes than one would get from the normal distribution with the same mean and variance
- Also implies that there is a lower probability of non-extreme outcomes
- Probabilities are between zero and one, so the area under the distribution is one

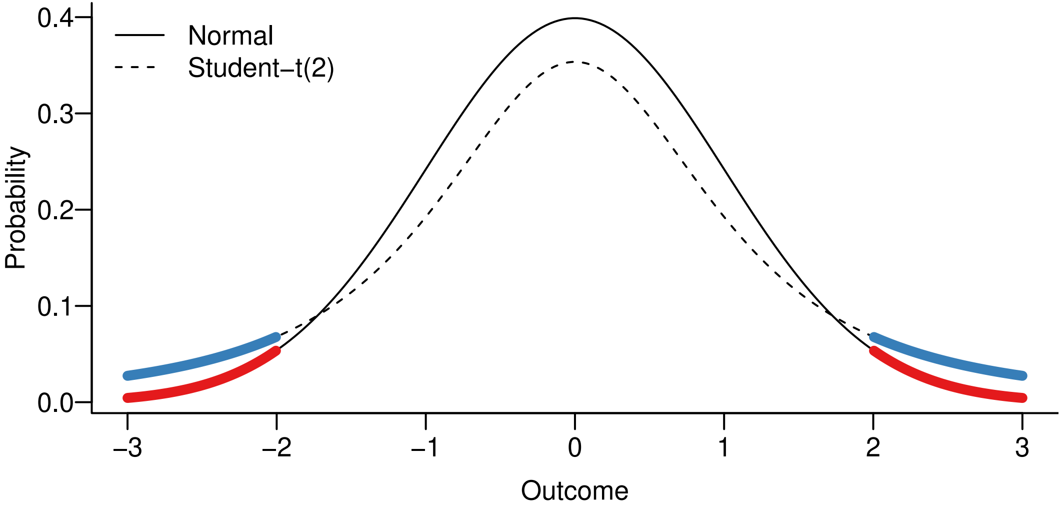
The Student-t Distribution

- The degrees of freedom – ν – of the Student-t distribution indicate how fat the tails are
- $\nu = \infty$ implies the normal
- $\nu < 2$ superfat tails
- For a typical stock $3 < \nu < 5$
- The Student-t is convenient when we need a fat-tailed distribution

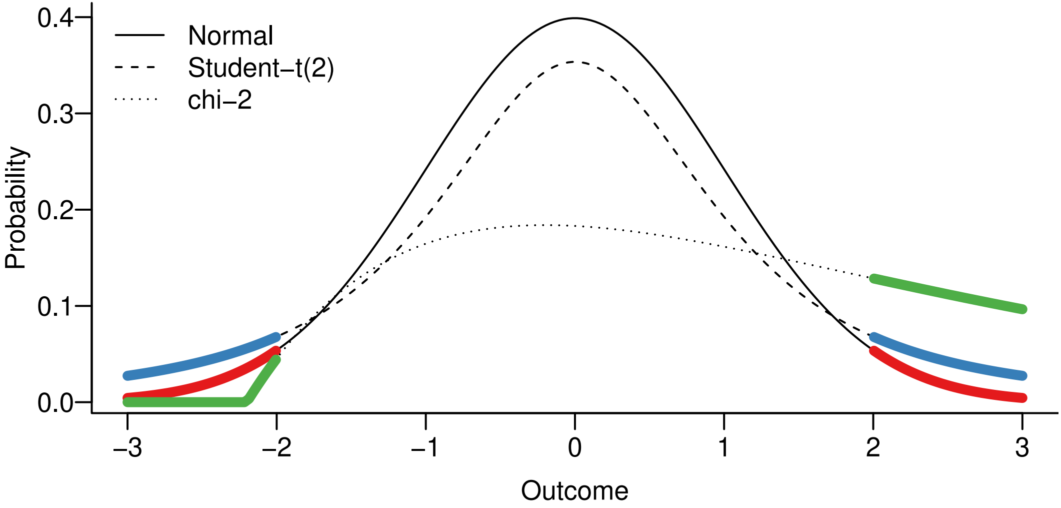
Tails

Tails

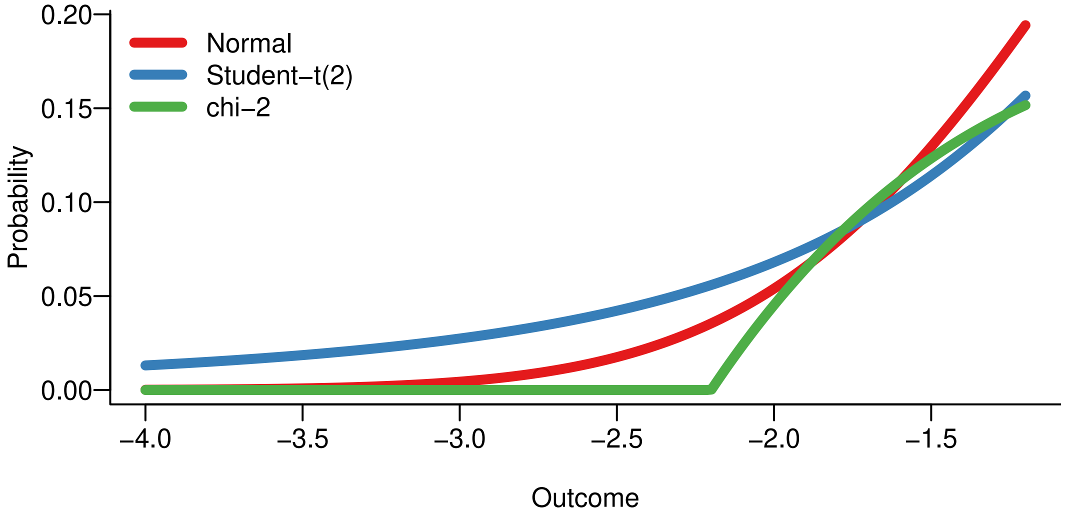
Tails


Normal
Student–t(2)
Probability
Outcome
0.0
0.1
0.2
0.3
0.4
–3
–2
–1
0
1
2
3

Tails


Normal
Student–t(2)
chi–2
Probability
Outcome
0.0
0.1
0.2
0.3
0.4
–3
–2
–1
0
1
2
3

Zoom Into The Tails

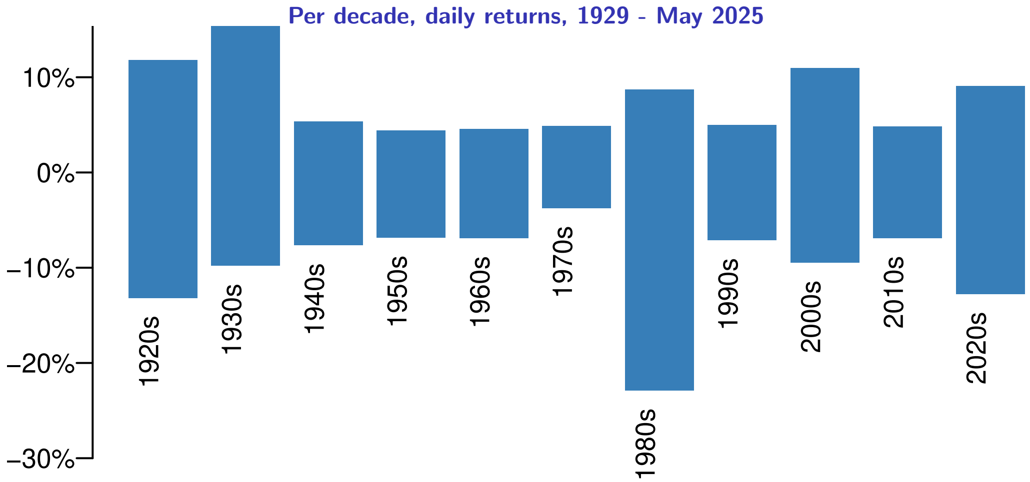


Probability of Extreme Outcomes

- If S&P-500 returns were normally distributed, the probability of a one-day drop of 23% would be 3×10^{-89}
  in R: `pnorm(-0.23,sd=0.01151996)= 5.512956e-89`
- The table below gives probabilities of different returns assuming normality

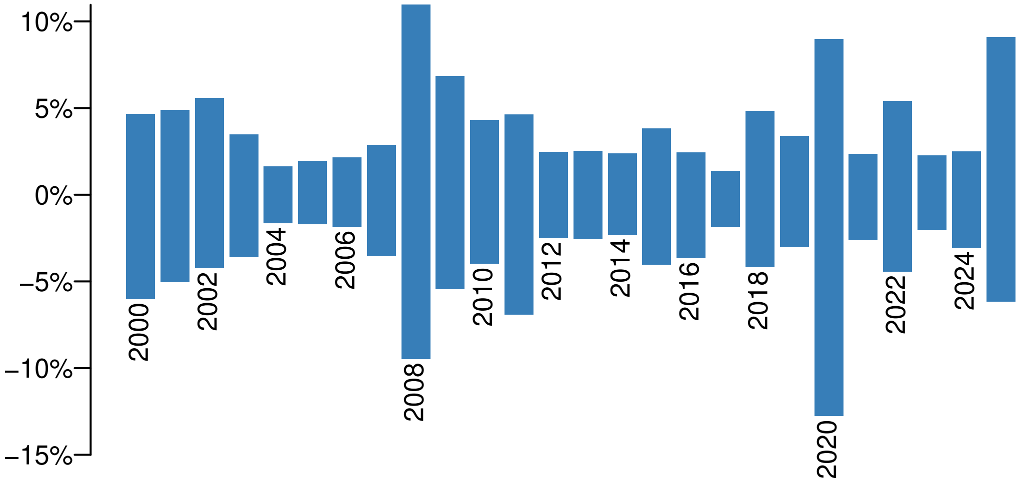
Returns above or below	Probability
1%	0.385
2%	0.0820
3%	0.00909
5%	1.37×10^{-5}
15%	6.92×10^{-39}
23%	5.51×10^{-89}

Maximum and Minimum of S&P-500 Returns



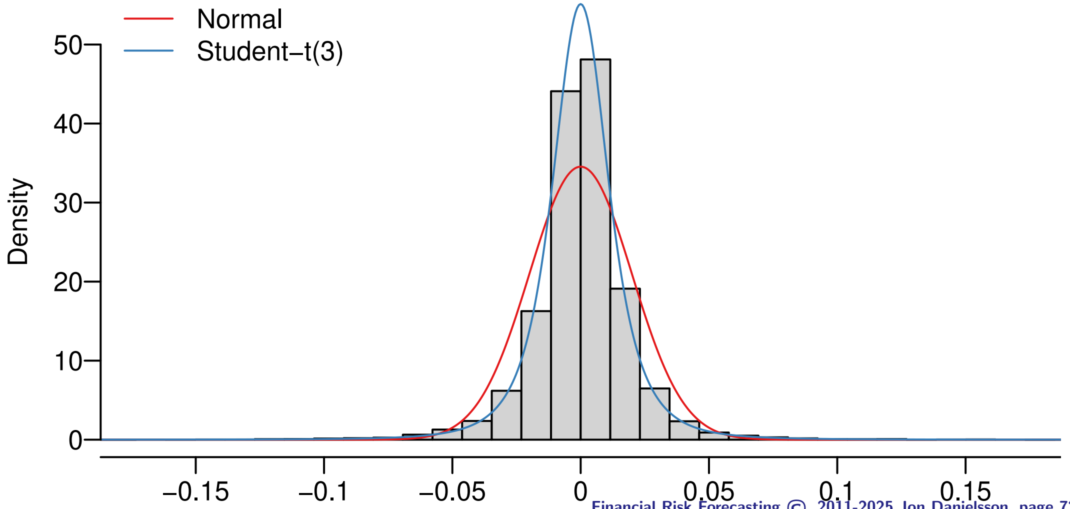
Max and Min of S&P-500 Returns

Per year, daily returns, 2000 - May 2025

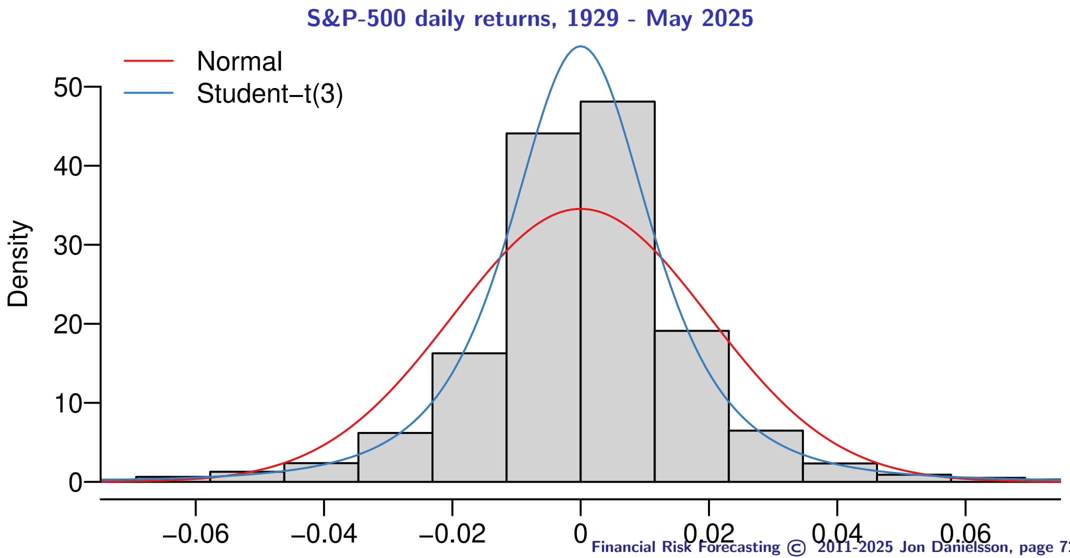


Empirical Density vs Normal and t(3)

S&P-500 daily returns, 1929 - May 2025

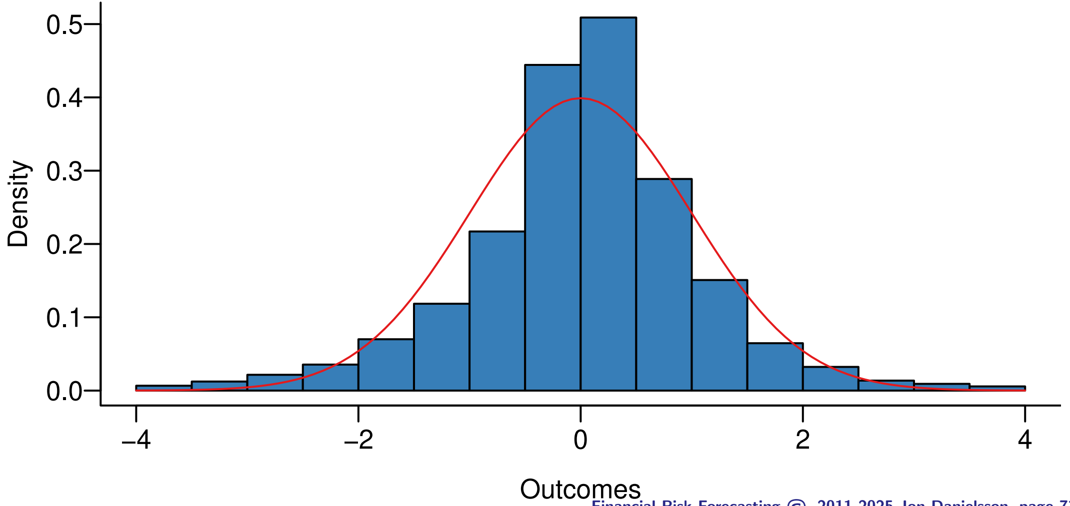


Zoomed Empirical Density vs Normal and t(3)



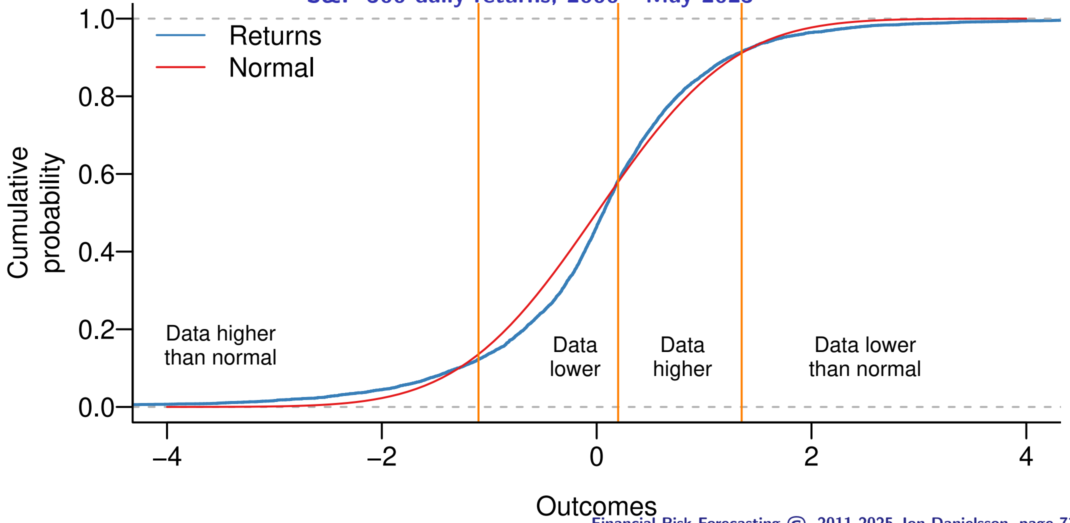
Empirical Density vs Normal

S&P-500 daily returns, 2000 - May 2025



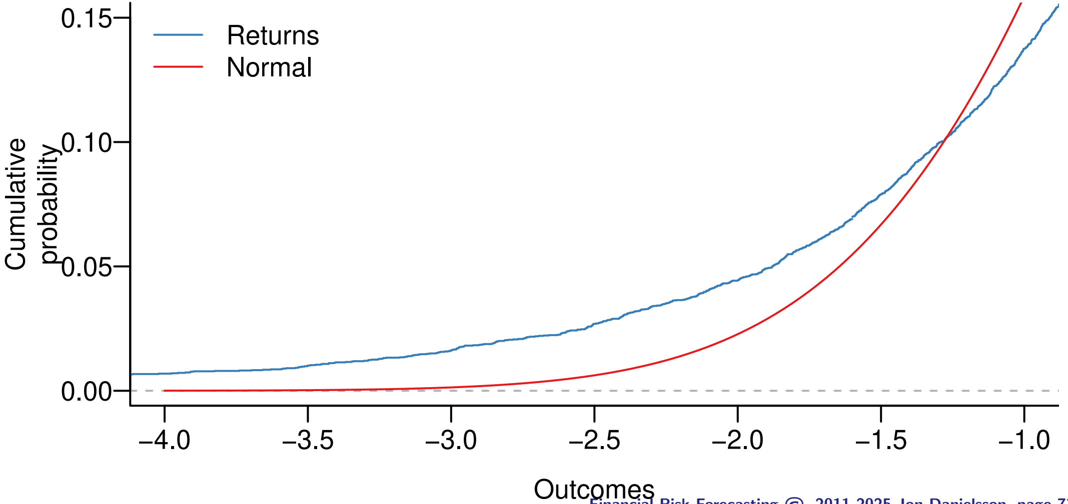
Empirical Density vs Normal

S&P-500 daily returns, 2000 - May 2025



Empirical Density vs Normal

S&P-500 daily returns, 2000 - May 2025



Non-Normality and Fat Tails

- Three observations
  1. Peak is higher than normal
  2. Sides are lower than normal
  3. Tails are much thicker (fatter) than normal

Identifying Fat Tails

Identification of Fat Tails

- Two main approaches for identifying and analysing tails of financial returns: statistical tests and graphical methods
- The *Jarque-Bera* (JB) and the *Kolmogorov-Smirnov* (KS) tests can be used to test for fat tails
- QQ plots allow us to analyse tails graphically by comparing quantiles of sample data with quantiles of reference distribution

Jarque-Bera Test

- The Jarque-Bera (JB) test is a test for normality and may point to fat tails if rejected
- The JB test statistic is:

$$\frac{T}{6}\text{Skewness}^2 + \frac{T}{24}(\text{Kurtosis} - 3)^2 \sim \chi^2_{(2)}$$

R

```
library(tseries)
jarque.bera.test(y)
```

Kolmogorov-Smirnov Test

- Based on minimum distance estimation comparing sample with a reference distribution, like the normal

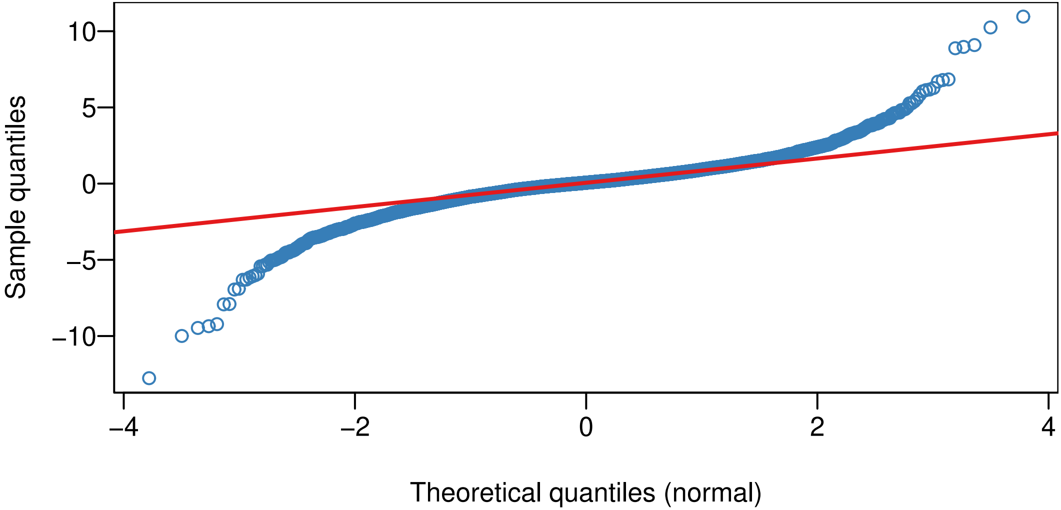
QQ Plots

- A QQ plot (quantile-quantile plot) compares the quantiles of sample data against quantiles of a reference distribution, like normal
- Used to assess whether a set of observations has a particular distribution
- Can also be used to determine whether two datasets have the same distribution
- The x-axis show quantiles from a standard distribution (like $\mathcal{N}(0,1)$)
- The y-axis show what values would be expected if data followed same distribution but with a different standard deviation (line) and what data actually is (dots)

R

```
library(car)
qqPlot(y)
qqPlot(y, distribution="t", df=5)
```

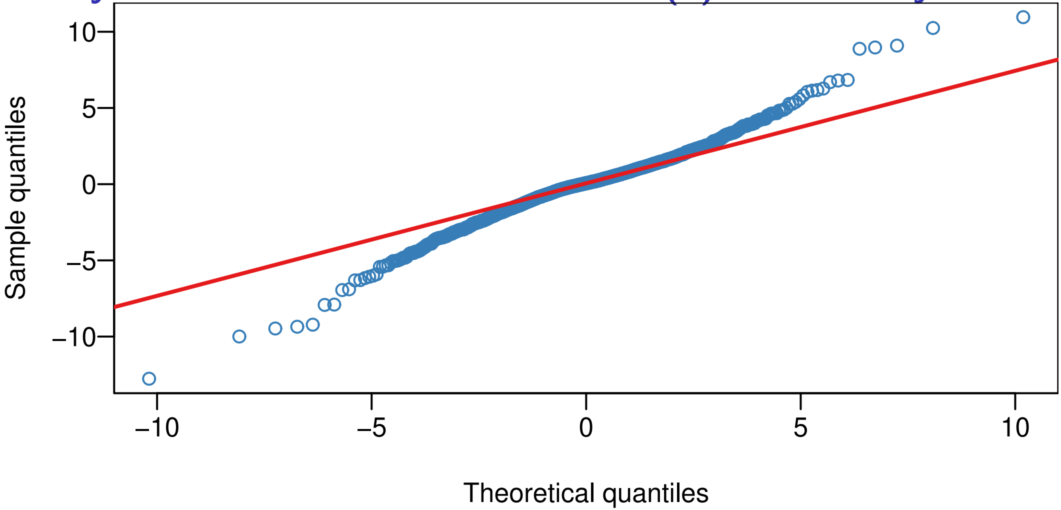
Daily S&P-500 Returns vs Normal: 2000 - May 2025



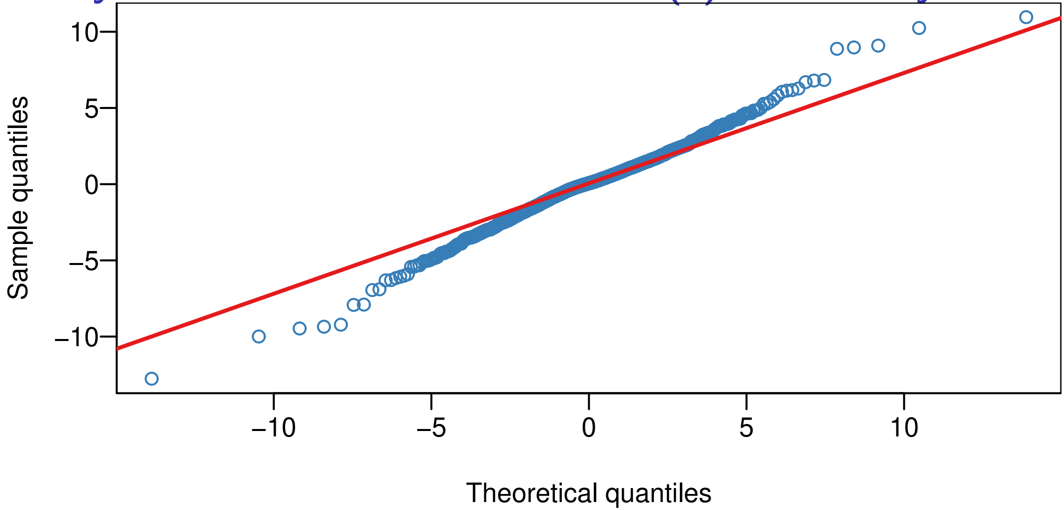
Daily S&P-500 Returns vs Normal

- Many observations seem to deviate from normality and the QQ-plot has clear S shape
- Indicates that returns have fatter tails than normal, but how much fatter?
- We can use the Student-t with different degrees of freedom as reference distribution (fewer degrees of freedom give fatter tails)

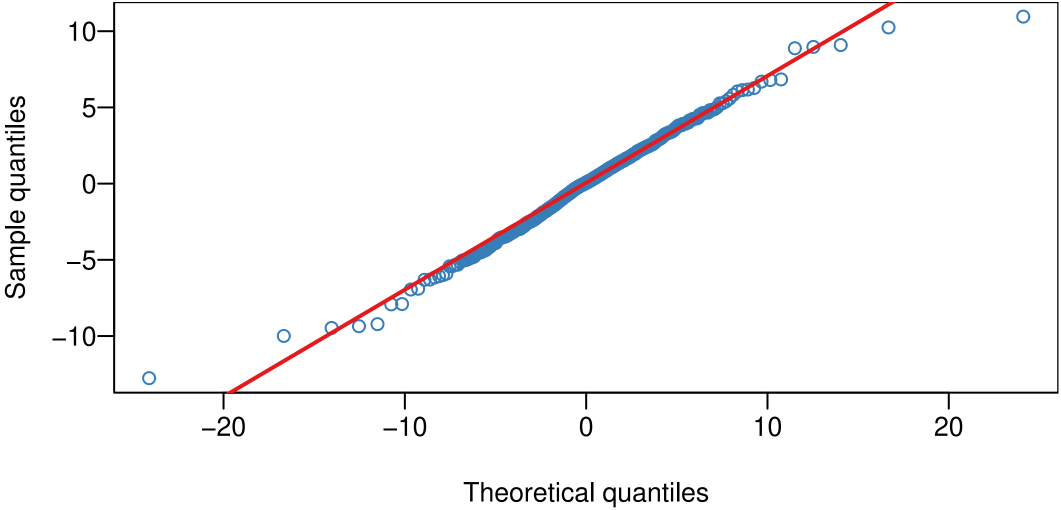
Daily S&P-500 Returns vs Student-t(5) 2000 - May 2025



Daily S&P-500 Returns vs Student-t(4) 2000 - May 2025



Daily S&P-500 Returns vs Student-t(3) 2000 - May 2025



Non-linear Dependence

Correlations

- Correlations are a linear concept

$$y = \alpha x + \epsilon$$

- Then α is proportional to the correlation between x and y
- A different way to say that is *linear dependence*
- The relationship between the two variables is always the same regardless of the magnitude of the variables
- Under the normal distribution, dependence is linear
- Key assumption for the mean–variance model

Non-linear Dependence

- Non-linear dependence (NLD) implies that dependence between variables changes depending on some factor, in finance, perhaps according to market conditions
  - Example: Different returns are relatively independent during normal times, but highly dependent during crises
- If returns were jointly normal, correlations would decrease for extreme events, but empirical evidence shows exactly the opposite
- Assumption of linear dependence does not hold in general

Evidence of Non-linear Dependence

Daily returns for Microsoft, Morgan Stanley, Goldman Sachs and Citigroup

5 May 1999 - 12 June 2015

	MSFT	MS	GS
MS	46%		
GS	46%	81%	
C	37%	65%	63%

1 August 2007 - 15 August 2007

	MSFT	MS	GS
MS	93%		
GS	82%	94%	
C	87%	93%	92%

More on NLD

- We will return to NLD in Chapter 3

Issues with Volatility, Fat Tails and Nonlinear Dependence

Implications of NLD And Fat Tails

- Non-normality and fat tails have important consequences in finance
- Assumption of normality may lead to a gross underestimation of risk
- However, the use of non-normal techniques is highly complicated and unless correctly used may lead to incorrect outcomes

Volatility and Fat Tails

- Volatility is a correct measure of risk *if and only if* the returns are normal
- If they follow the Student-t or any of the fats, then volatility will only be partially correct as a risk measure
- We discuss this in more detail in Chapter 4

The Quant Crisis of 2007

- Many hedge funds using quantitative trading strategies ran into serious difficulties in June 2007
- The correlations in their assets increased very sharply
- So they were unable to get rid of risk

Goldman Sachs's Flagship Global Alpha Fund (Summer of 2007)

"We were seeing things that were 25-standard deviation moves, several days in a row," said David Viniar, Goldman's chief financial officer. "There have been issues in some of the other quantitative spaces. But nothing like what we saw last week."

Lehman Brothers (Summer of 2007)

"Wednesday is the type of day people will remember in quantland for a very long time," said mr. Rothman, a University of Chicago PhD, who ran a quantitative fund before joining Lehman Brothers. "Events that models only predicted would happen once in 10,000 years happened every day for three days."

Volatility and Fat Tails

- Goldman's 25-sigma event under the normal has a probability of 3×10^{-138}
- Age of the universe is estimated to be 5×10^{12} days while the earth is 1.6×10^{12} days old
- Goldman expected to suffer a one-day loss of this magnitude less than one every 1.5×10^{125} universes
- Or perhaps the distributions were really not Gaussian

Diversification and Fat Tails

- Suppose you go to a dodgy buffet restaurant
  - Where you worry about food poisoning in one of the foods offered
  - But you don't know which
  - And are really hungry
  - How many different types of food do you try?
- When the tails are super fat, diversification may not be advisable

R and Data

Implementing empirical techniques

- We have three general choices: Excel, general-purpose programmes like Stata or a statistical programming language
- Matlab, Python, Julia, R
- In this course we pick R for three reasons
  1. it provides the best user interface (RStudio)
  2. it is easiest to get started with it
  3. it is generally best for statistical work
- The R notebook provides a comprehensive introduction to R as used in this course
- A part of the weekly classes is dedicated to R
- And LLMs are very useful for learning and implementing code

Data

- It is difficult to get high quality financial data in an easily accessible way
- You can use Bloomberg or WRDS, but they are very complicated
- There are several vendors of financial data
- finance.yahoo.com is often used and is free
- We suggest a vendor called EOD — Free access is provided to LSE students in this course
- https://eodhd.com/financial-academy/
- We will demonstrate R and EOD in classes and now quickly demonstrate them

R and RStudio

- R is a powerful open-source language for statistics, data analysis and simulation
- It is widely used in finance, economics and academia
- RStudio is a user-friendly free to use interface for R:
  - Makes it easier to write, test and debug code
  - Includes built-in tools for plots, packages and version control
- All examples in this course use R — including forecasting, simulation and plotting
- You can install both from: posit.co/download/
- We use scripts and notebooks — not the console alone

Using eodhdR2 to Access Financial Data in R

- The `eodhdR2` package provides a simple interface to the EOD Historical Data API
- https://eodhd.com/financial-apis/r-library-v-2-for-financial-data-by-eodhd-2024
- It relies on several supporting packages:
  - `httr` — handles HTTP requests to the API
  - `jsonlite` — parses JSON responses from the server
  - `readr` — reads CSV-formatted data (if used)
  - `lubridate` — parses and formats dates
  - `data.table` — efficient handling of time series data
- Once loaded, you can download and analyse financial data directly from R

Example 1 — Setup

```
install.packages(c("httr", "jsonlite", "lubridate",
    "data.table", "readr"))
install.packages('eodhdR2')
library(eodhdR2)
token = "demo"
token <- "demo"
set_token(token)
help(get_prices)
```

Example 2 — Get Data and Plotting

```
get_dividends("AAPL", "US")
prices = get_prices("AAPL", "US")
head(prices)
plot(prices$date,prices$close,type='l')
splits =  get_splits("AAPL", "US")
 # Error in `get_splits()`:
# You need a proper token (not demonstration) for exchange list..
```

Example 3 — More Plot

```
plot(prices$date, prices$close,
    type = "l",
    col = "blue",
    main = "AAPL Closing Prices",
    xlab = "Date",
    ylab = "Close")
```

Simulations

The focus of Chapter 7

Idea

- Replicate a part of the world in computer software
- For example, market outcomes, based on some model of market evolution
- Sufficient number of simulations (replications) ideally yield a large and representative sample of market outcomes
- Use that to calculate quantities of interest, perhaps risk or performance

Obtaining Random Numbers

- The fundamental input in Monte Carlo (MC) analysis is a long sequence of random numbers (*RNs*)
- Creating a large sample of *high-quality* RNs is difficult
- It is impossible to obtain pure random numbers
  - There is no natural phenomena that is purely random
  - Computers are deterministic by definition
- A computer algorithm known as a *pseudo random number generator* (RNG) creates outcomes that *appear* to be random even if they are deterministic

Random Numbers in R

```
runif(n=1)
runif(n=1,min=0,max=10)
rnorm(n=1)
rnorm(n=1,mean=-10,sd=4)
rt(n=1,df=4)
rnorm(n=100)
```

S&P-500 2015 to April 2024

- The unconditional volatility is 1.15%
- Unconditional mean 0.038%
- What might happen over the next day, month, year and decade
- Simulate a random walk

Simulate a Random Walk

- Start with a price, p_t and a distribution of returns
- And simple (arithmetic) returns (could have used log returns)

$$r_t \sim \mathcal{N}(\mu, \sigma^2)$$

$$r_t \sim \mathcal{N}(0.00038, 0.01146^2)$$

- Want to simulate one day into future (decorate simulations with a tilde)
- Call simulated return $\tilde{r}_{t+1}$
- The simulated tomorrow price is then:

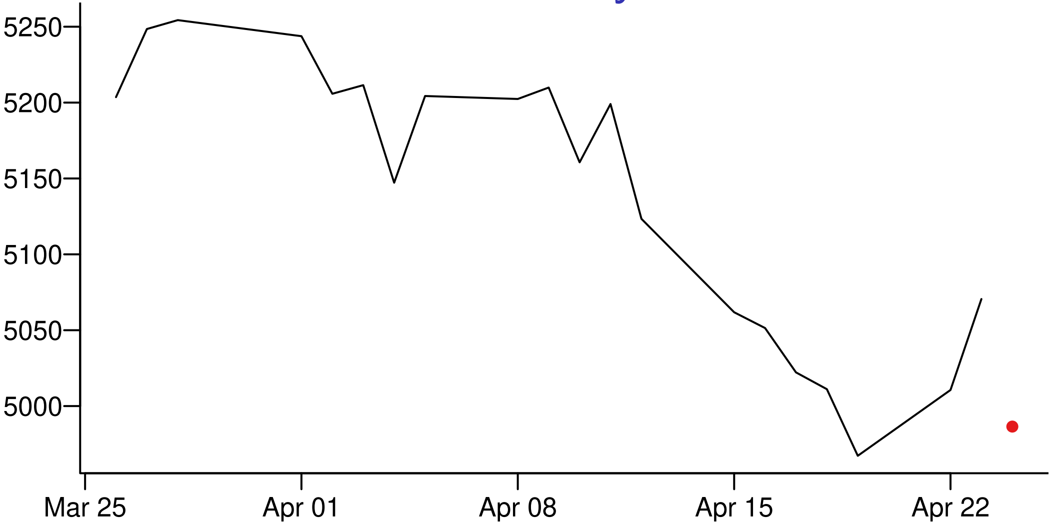
$$\tilde{p}_{t+1} = p_t(1 + \tilde{r}_{t+1})$$

Random Walk in R

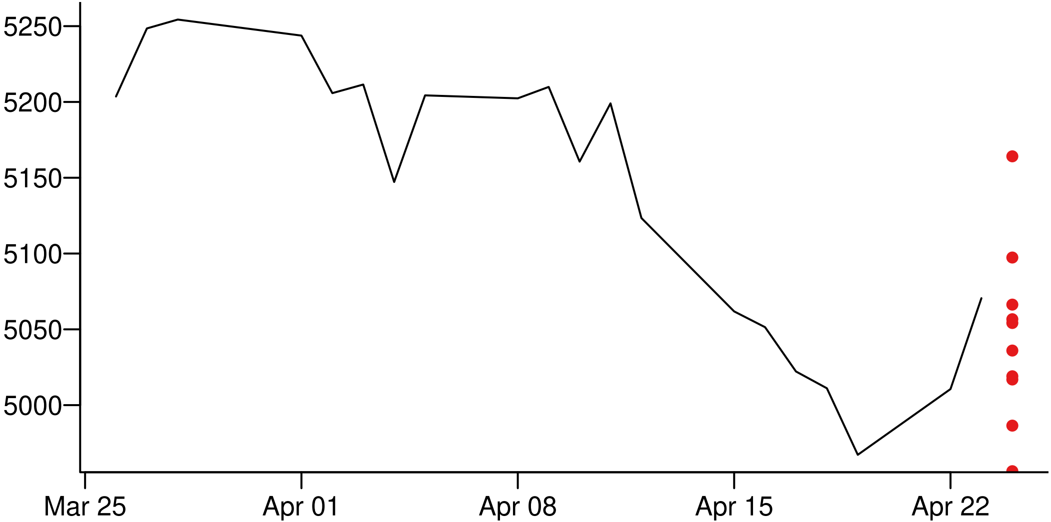
Set seed, simulate returns, cumulative product (`cumprod`), normalise to start at one, multiply by last price

```
set.seed(seed)

simR=rnorm(1,sd=sigma,mean=mu+1)
simP=P * (1+simR)

simP=cumprod(simR)
simP=simP/simP[1]
simP=simP*Price
```

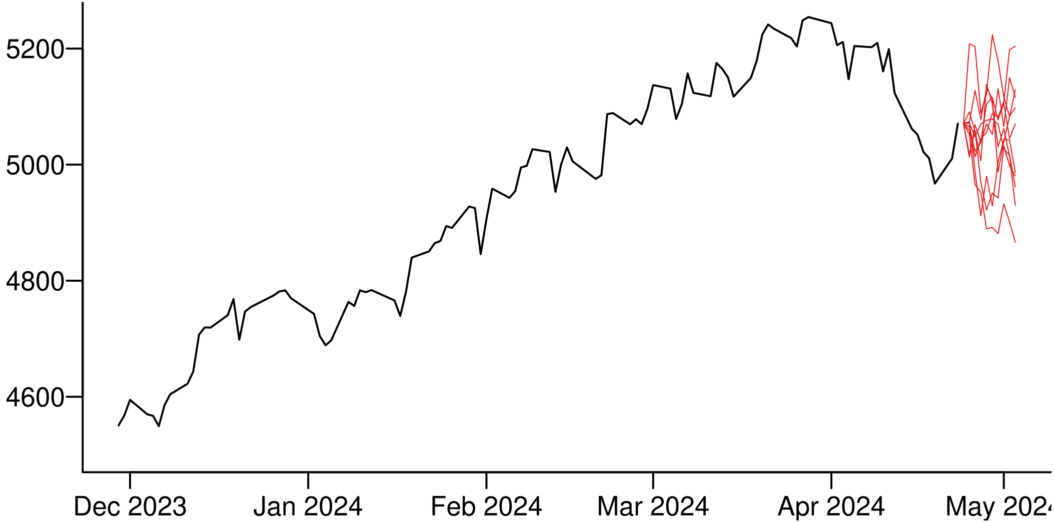
S&P-500, One day 1 sim



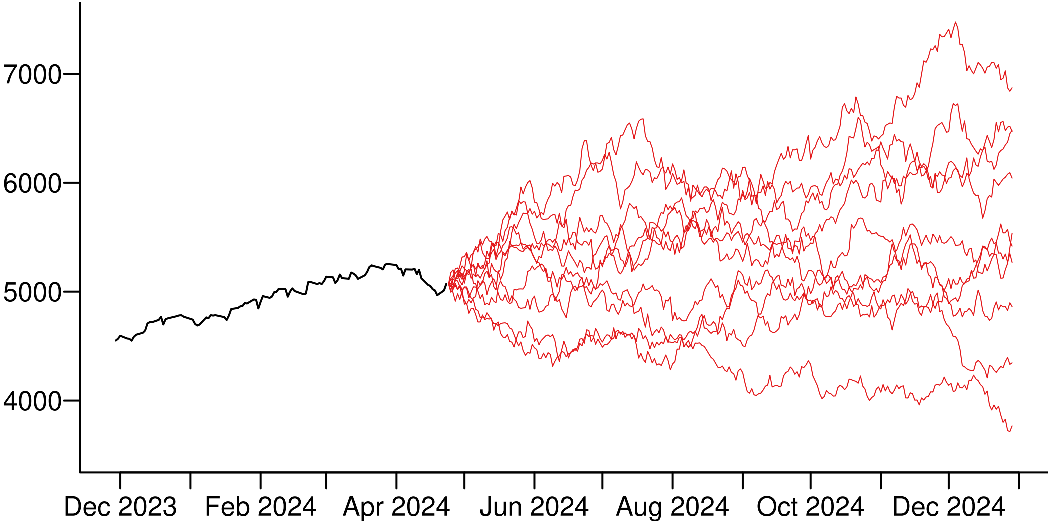
S&P-500, One day 8 sims



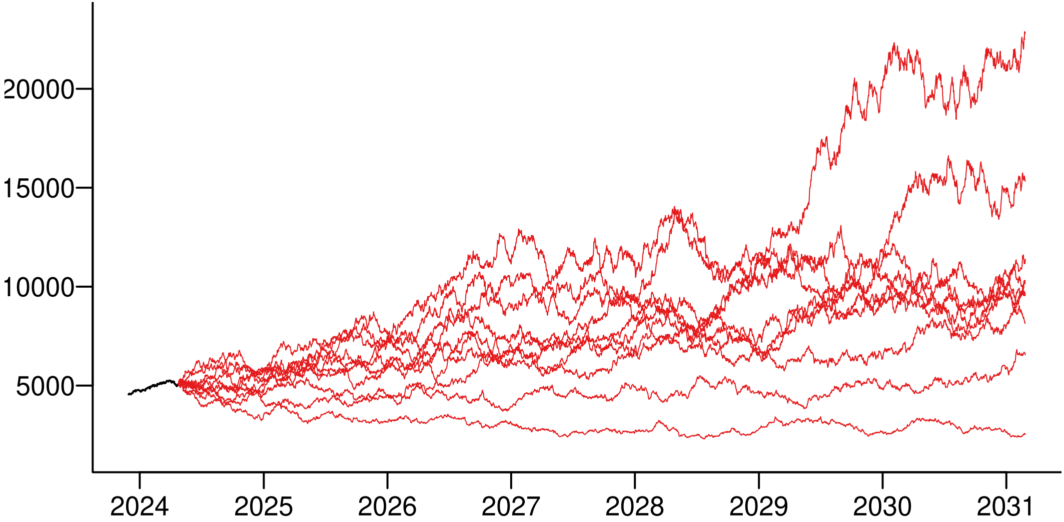
S&P-500, More Days



S&P-500, More Days



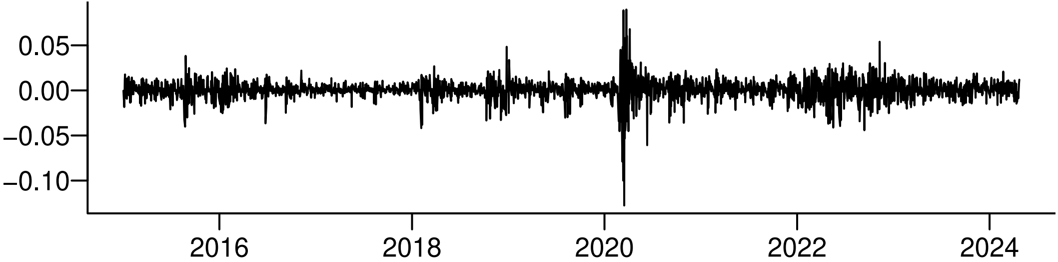
S&P-500, More Days



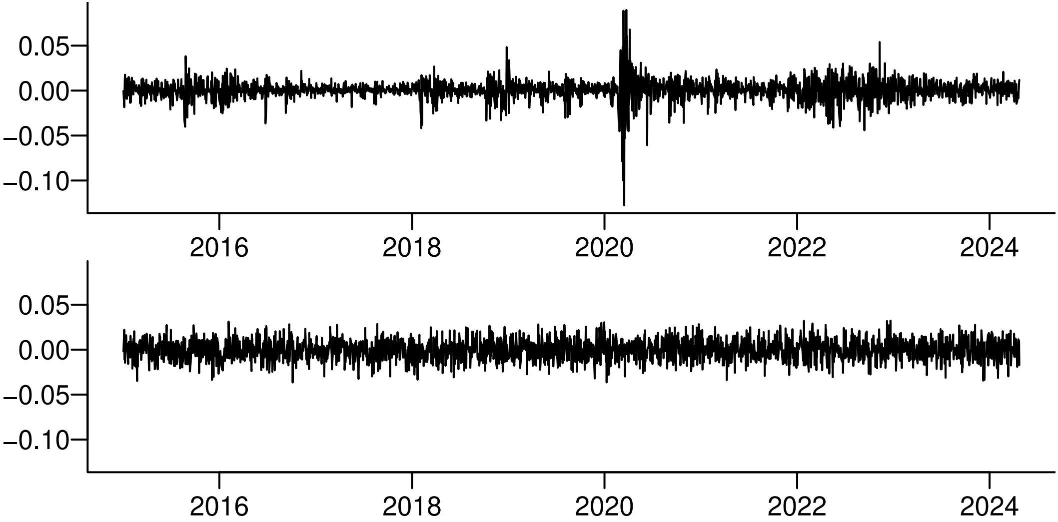
Summary

- Some of the simulated prices seem quite unreasonable
- May be better to look at the returns

S&P-500 Returns



S&P-500 Returns with one simulated path



Summary

- The actual S&P-500 exhibits a number of volatility clusters
- For example, March 2020, end of year 2018, 2017 and second part of 2015
- The simulated returns look quite different

Applications of Risk Forecasting

What we do

- Quantitative methods for forecasting risk
- The underlying technology has many applications beside risk
- Such as in the management of investment portfolios
- Price forecasting and hence trading

Internal

- Every financial institution needs to manage risk and that means using quantitative techniques of the type we see in this course
- They are both used for managing risk and also to forecast risk and trading
- Some develop them in-house
- Others buy them in

Regulations and outside the restrictions

- Every financial institution is regulated
- Banks with the Basel Accords (see Chapter 10)
- Hedge funds and other investment managers are subject to mandates that usually include risk as a core component
- A very extensive and growing need for compliance further increases the need for quantitative techniques

RMaaS: Risk Management as a Service

- With so many IT functions moving into the cloud and hired *as a service* *aaS
- So has risk management
- Two main platforms *RiskMetrics* and Blackrock's *Aladdin* (much the bigger)
- Financial institutions can buy everything they need from Aladdin, up to all risk management and risk modelling
- Done automatically by Aladdin's AI

Machine Learning and Risk Forecasting

Machine Learning (ML)

- Machine learning aims to estimate a function f such that

$$y \approx f(x)$$

  where x is input data and y is the predicted output
- Neural networks, decision trees and support vector machines are common ML tools
- ML learns patterns directly from data — without explicit model assumptions
- Typically trained by minimising prediction error (e.g. mean squared error)
- Often used in applications like image recognition, credit scoring and high-frequency trading
- No assumptions about distributions, but requires large, representative datasets

ML vs Traditional Statistical Modelling

	Traditional Statistics	Machine Learning
Goal	Parameter estimation, inference	Prediction accuracy
Data requirements	Can work with small datasets	Needs large datasets
Model structure	Specified in advance (e.g. GARCH, ARIMA)	Learned from data (e.g. neural net)
Assumptions	Explicit (e.g. normality, independence)	Few or none
Interpretability	High (parameters have meaning)	Often low (black box)
Use in risk forecasting	Common, robust, well-understood	Less reliable, high variance

Why Not Use ML for Risk Forecasting?

- You might ask: why not use something like PyTorch or TensorFlow here?
- ML methods require large datasets to accurately learn patterns
- In risk forecasting, we often work with limited data — a few thousand daily returns
- More importantly, we already have strong *prior knowledge* about market dynamics
- Traditional models are designed to reflect this structure (e.g. GARCH) with perhaps 3-4 parameters estimated from modest data
- ML typically requires hundreds or thousands of parameters — which increases risk of overfitting
- ML models often lack interpretability and are harder to validate under regulatory scrutiny
- But for forecasting tail risk, transparency and tractability matter more than raw prediction power

Hybrid Approaches: ML and Traditional Models Together

- ML is not a replacement for traditional risk models — but it can enhance them
- In a hybrid approach, ML is used to:
  - Monitor model performance over time
  - Detect structural breaks or regime changes
  - Select between candidate models based on predictive accuracy
  - Identify nonlinear features or interactions in pre-processing
- The traditional model still produces the risk forecast
- This balances the interpretability of statistics with the adaptability of ML

AI in Risk Forecasting: Human-on-the-Loop

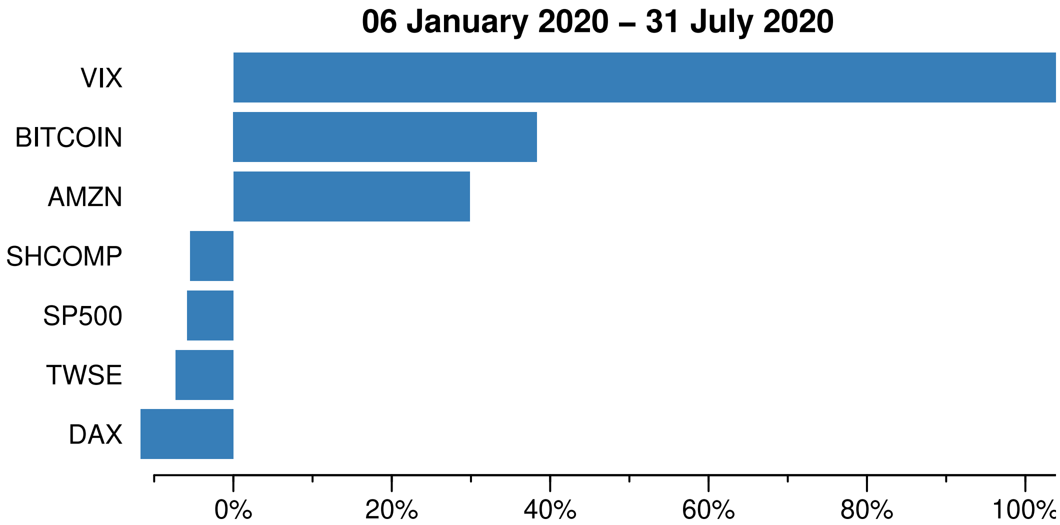
- In this model, AI handles routine tasks, while the human monitors and intervenes when necessary
- Relevant to our course:
  - AI automates model fitting across many assets
  - Flags poor performance or structural breaks
  - Suggests alternative specifications (e.g. GARCH(1,1) vs apARCH)
- The human evaluates the suggestions, checks diagnostics and makes final modelling decisions
- This setup scales well and keeps expert judgement central
- Already used in institutional risk monitoring and model governance frameworks

Where Human-on-the-Loop AI Is Used in Practice

- BlackRock — Aladdin
  - Combines statistical risk models with AI-based anomaly detection
  - Human risk managers oversee and approve alerts and overrides
- MSCI RiskMetrics
  - Uses traditional VaR/ES modelling enhanced with machine learning for portfolio-wide diagnostics
  - Human analysts validate outliers and model breaks
- JP Morgan — LOXM platform
  - Executes trades using reinforcement learning under human supervision
  - Similar techniques applied to internal risk monitoring tools
- Regulatory context
  - Banks are expected to keep humans in control of model changes under Basel governance rules
  - "Explainability" and human validation are core to model risk management

Covid-19

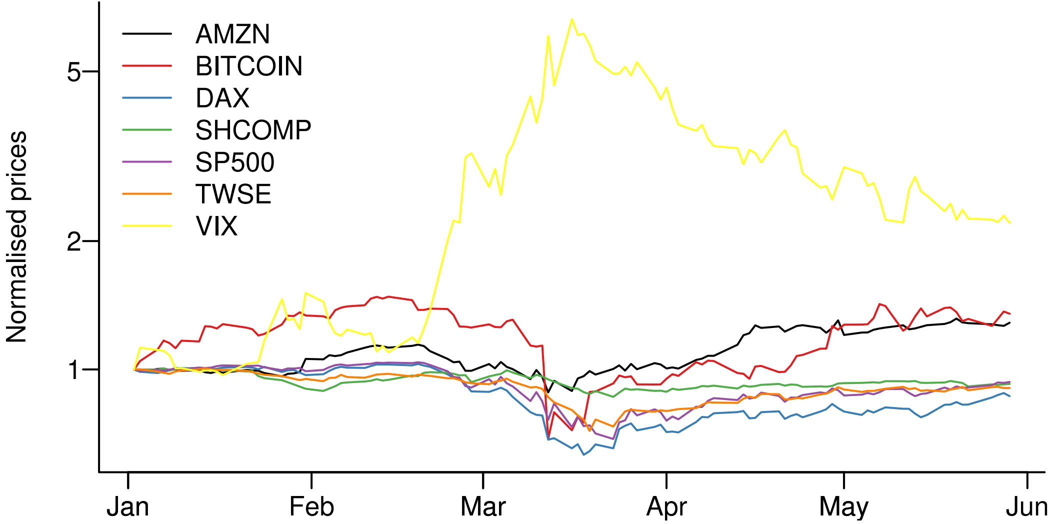
Market Performance



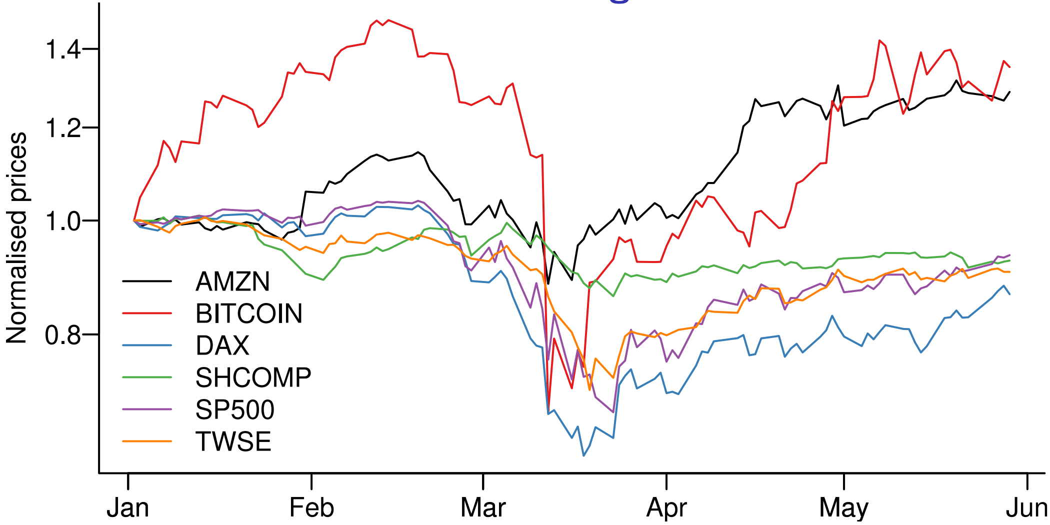
Normalised Prices

- By normalising the price of each of the assets to one at the beginning of 2020, we can see how they performed throughout the crisis
- The most remarkable are Bitcoin and SHCOMP

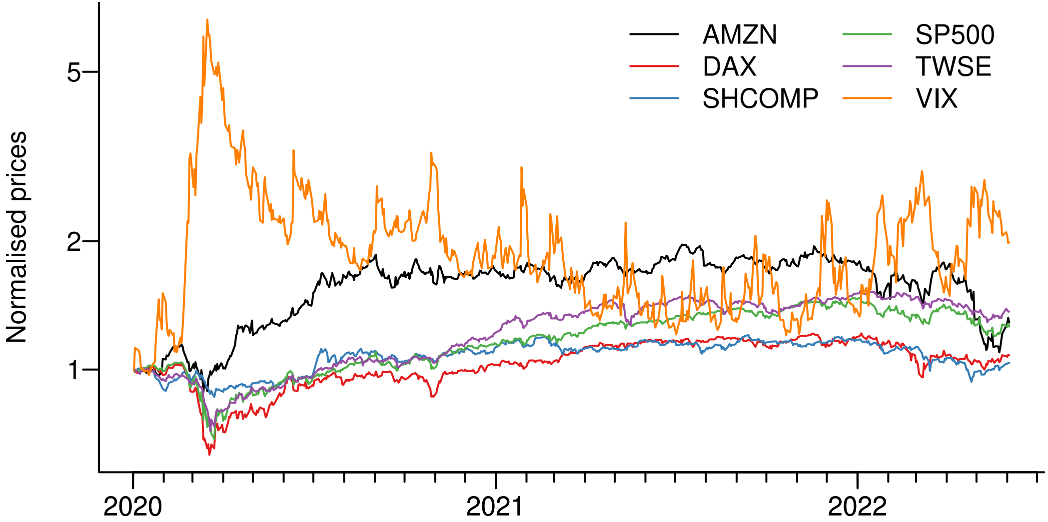
Normalised Prices



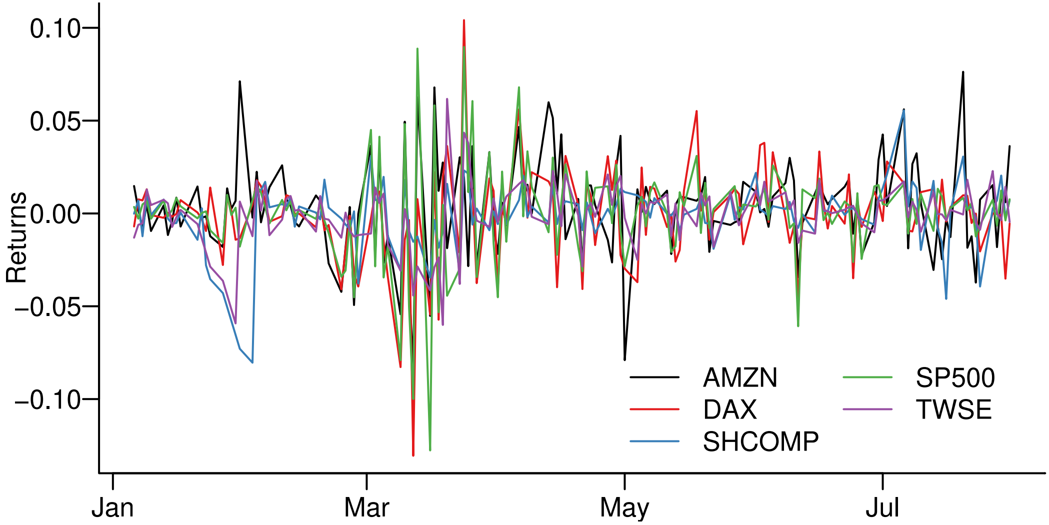
Normalised Prices. Zooming Into Main Crisis



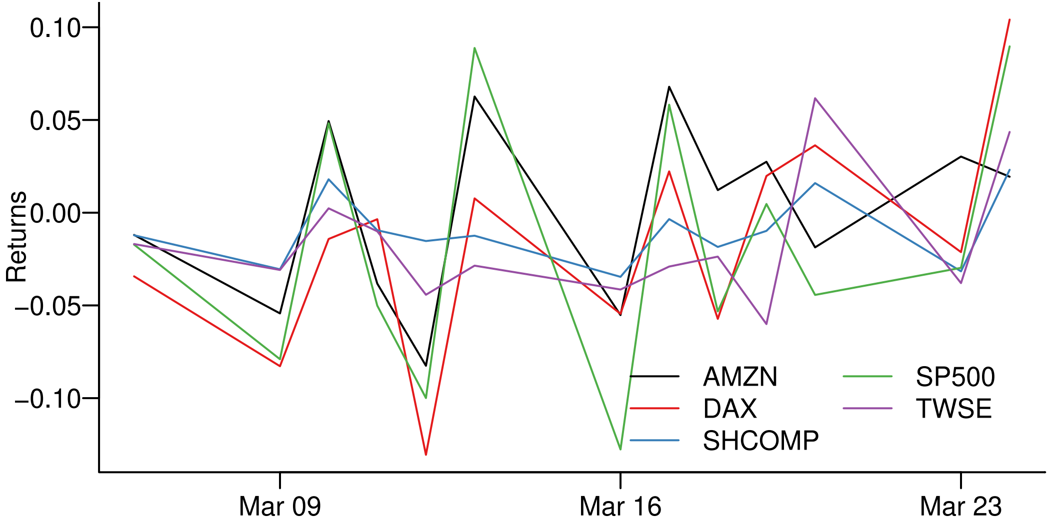
Normalised Prices



Returns


0.10
0.05
0.00
−0.05
−0.10
Returns
Jan
Mar
May
Jul
AMZN
DAX
SHCOMP
SP500
TWSE

Returns: Zooming Into Main Crisis



Returns

- It is hard to see much with the return plots
- The US dollar-euro exchange rate is the most stable and Bitcoin the least stable
- And we see a clear volatility cluster in March
- And we will consider that in much more detail later

Non-linear Dependence

- By looking at correlations between returns in the full sample and at the height of the crisis
  - The crisis correlations are much higher
- A clear example of non-linear dependence
- In turn, any volatility model will need to pick that up as we see in Chapter 3

S&P-500

- The S&P-500 is the asset we spend most of the time in this course on
- And while it certainly shows the impact of the crisis
- What is interesting is how little it is affected by the crisis

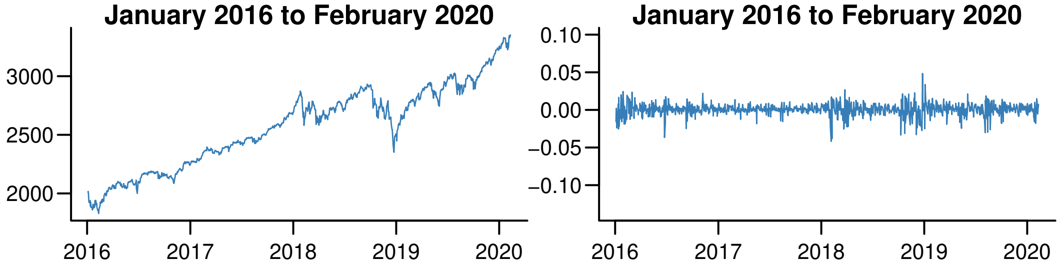
Correlations

- January 2020 to July 2021
- March 2020

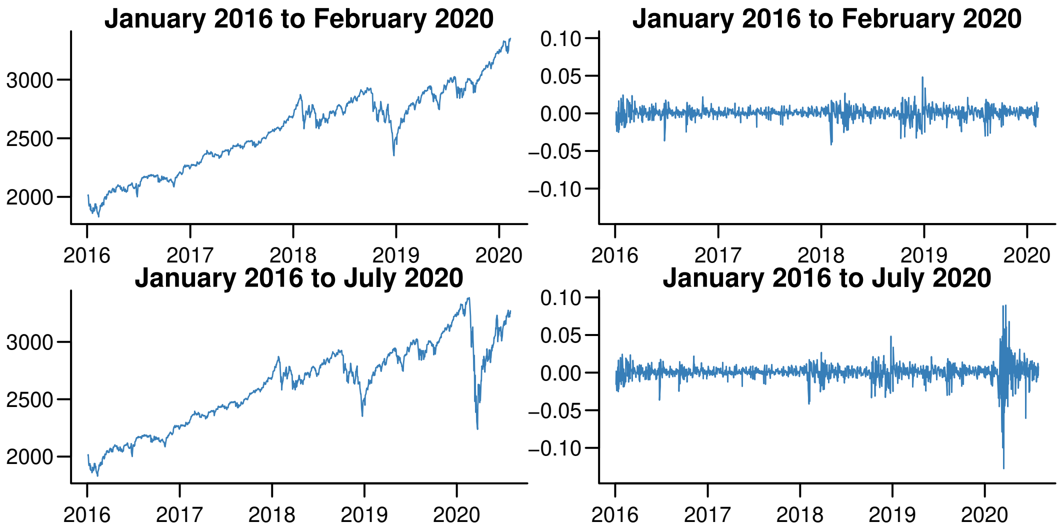
	AMZN	BITCOIN	DAX	SHCOMP	SP500	TWSE
AMZN	1.00	0.39	0.38	0.28	0.64	0.12
BITCOIN	0.39	1.00	0.47	0.11	0.47	0.09
DAX	0.38	0.47	1.00	0.36	0.71	0.52
SHCOMP	0.28	0.11	0.36	1.00	0.29	0.53
SP500	0.64	0.47	0.71	0.29	1.00	0.31
TWSE	0.12	0.09	0.52	0.53	0.31	1.00

	AMZN	BITCOIN	DAX	SHCOMP	SP500	TWSE
AMZN	1.00	0.71	0.61	0.37	0.86	0.06
BITCOIN	0.71	1.00	0.68	0.12	0.57	0.10
DAX	0.61	0.68	1.00	0.68	0.74	0.60
SHCOMP	0.37	0.12	0.68	1.00	0.63	0.78
SP500	0.86	0.57	0.74	0.63	1.00	0.29
TWSE	0.06	0.10	0.60	0.78	0.29	1.00

S&P-500


January 2016 to February 2020
3000
2500
2000
2016
2017
2018
2019
2020
January 2016 to February 2020
0.10
0.05
0.00
−0.05
−0.10
2016
2017
2018
2019
2020

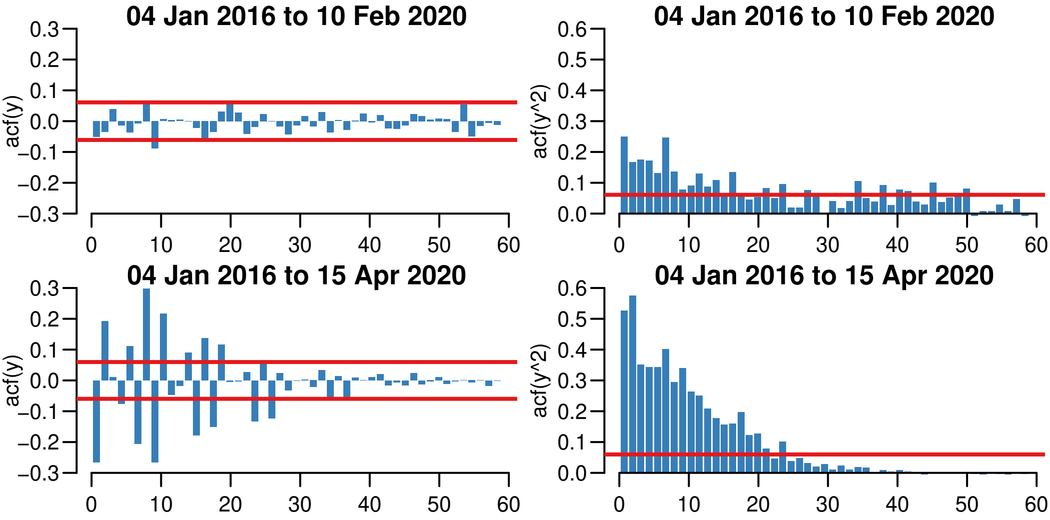
S&P-500


January 2016 to February 2020
3000
2500
2000
2016
2017
2018
2019
2020
January 2016 to February 2020
0.10
0.05
0.00
−0.05
−0.10
2016
2017
2018
2019
2020
January 2016 to July 2020
3000
2500
2000
2016
2017
2018
2019
2020
January 2016 to July 2020
0.10
0.05
0.00
−0.05
−0.10
2016
2017
2018
2019
2020

ACF Analysis of the S&P-500

- By showing the ACF of returns and return squared when we exclude and include 2020
- We see much stronger dependence in both the returns and volatility
- A question for you to consider is if the significant ACF implies violations of market efficiency
- And hence the ability to forecast the markets and hence make money

S&P-500 ACF With and Without 2020



Copulas

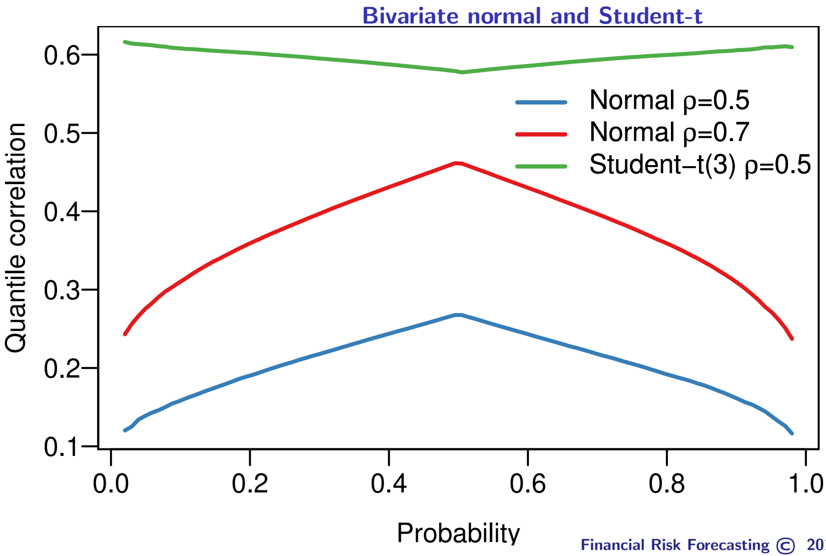
Exceedance Correlations

- Exceedance correlations show the correlations of (standardised) stock returns X and Y as being conditional on exceeding some threshold, that is,

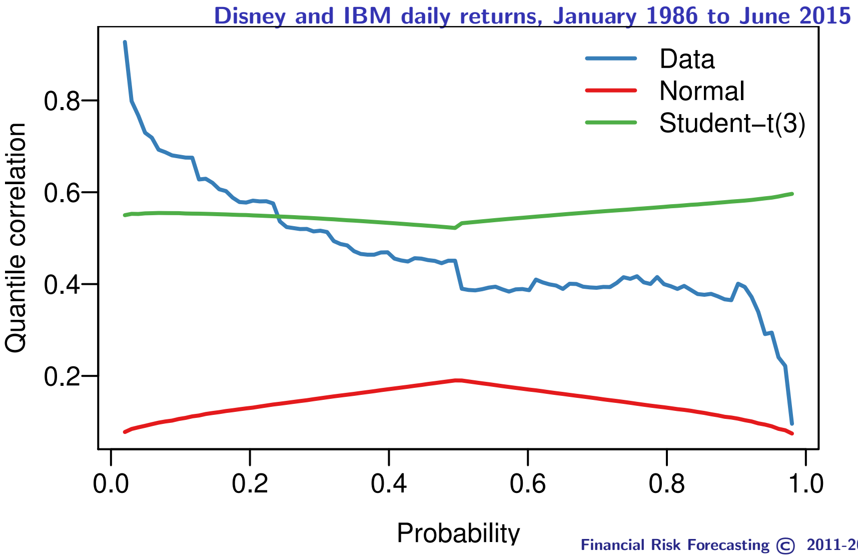
$$\tilde{\kappa}(p) = \begin{cases} \text{Corr}[X, Y | X \leq Q_X(p) \text{ and } Y \leq Q_Y(p)], & \text{for } p \leq 0.5 \\ \text{Corr}[X, Y | X > Q_X(p) \text{ and } Y > Q_Y(p)], & \text{for } p > 0.5 \end{cases}$$

  where $Q_X(p)$ and $Q_Y(p)$ are the p-th quantiles of X and Y given a distributional assumption
- Can be used to detect NLD

Exceedance Plot



Empirical Exceedance Plot



Copulas and Non-linear Dependence

- How do we model non-linear dependence more formally?
- One approach is multivariate volatility models (see Chapter 3)
- Alternatively we can use copulas, which allow us to create multivariate distributions with a range of types of dependence

Intuition Behind Copulas

- A copula is a convenient way to obtain the dependence structure between two or more random variables, taking NLD into account
- We start with the marginal distributions of each random variable and end up with a copula function
- The copula function joins the random variables into a single multivariate distribution by using their correlations

Intuition Behind Copulas

- The random variables are transformed to uniform distributions using the *probability integral transformation*
- The copula models the dependence structure between these uniforms
- Since the probability integral transform is invertible, the copula also describes the dependence between the original random variables

Theory of Copulas

- Suppose X and Y are two random variables representing returns of two different stocks, with densities f and g:

$$X \sim f \text{ and } Y \sim g$$

- Together, the joint distribution and marginal distributions are represented by the joint density h:

$$(X, Y) \sim h$$

- We focus separately on the marginal distributions (F, G) and the copula function C, which combines them into the joint distribution H

Theory of Copulas

- We want to transform X and Y into random variables that are distributed uniformly between 0 and 1, removing individual information from the bivariate density h

Theorem 1.1 Let a random variable X have a continuous distribution F and define a new random variable U as:

$$U = F(X)$$

Then, regardless of the original distribution F:

$$U \sim \text{Uniform}(0, 1)$$

Theory of Copulas

- Applying this transformation to X and Y we obtain:

$$U = F(X) \text{ and } V = G(Y)$$

- Using this we arrive at the following theorem

Theorem 1.2 Let F be the distribution of X, G the distribution of Y and H the joint distribution of (X,Y). Assume that F and G are continuous. Then there exists a unique copula C such that:

$$H(X, Y) = C(F(X), G(Y))$$

Theory of Copulas

- In applications we are more likely to use densities:

$$h(X, Y) = f(X) \times g(Y) \times C(F(X), G(Y))$$

- The copula contains all dependence information in the original density h, but none of the individual information
- Note that we can construct a joint distribution from any two marginal distributions and any copula and we can also extract the implied copula and marginal distributions from any joint distribution

The Gaussian Copula

- One example of a copula is the *Gaussian copula*
- Let $\Phi(\cdot)$ denote the normal (Gaussian) distribution and $\Phi^{-1}(\cdot)$ its inverse
- Let $U, V \in [0,1]$ be uniform random variables and $\Phi_\kappa(\cdot)$ the bivariate normal with correlation coefficient κ
- Then the Gaussian copula function can be written as:

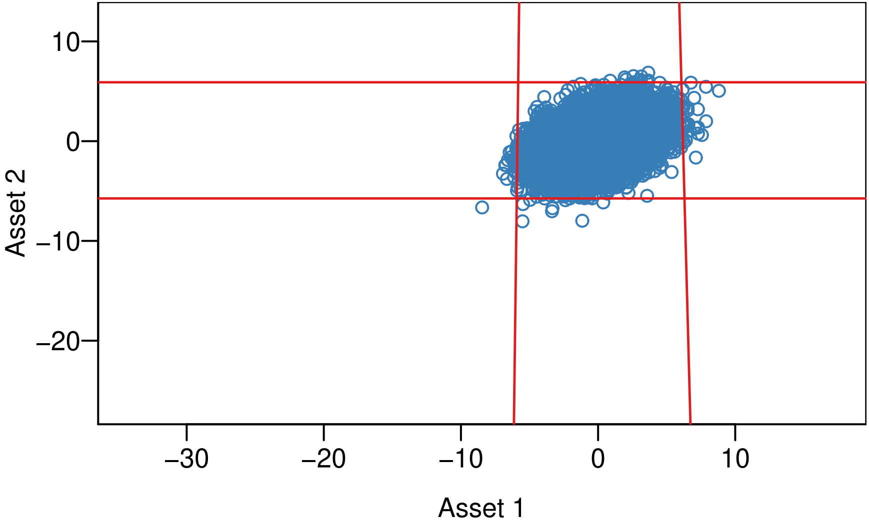
$$C(U,V) = \Phi_\kappa(\Phi^{-1}(U), \Phi^{-1}(V))$$

- This function allows us to join the two marginal distributions into a single bivariate distribution

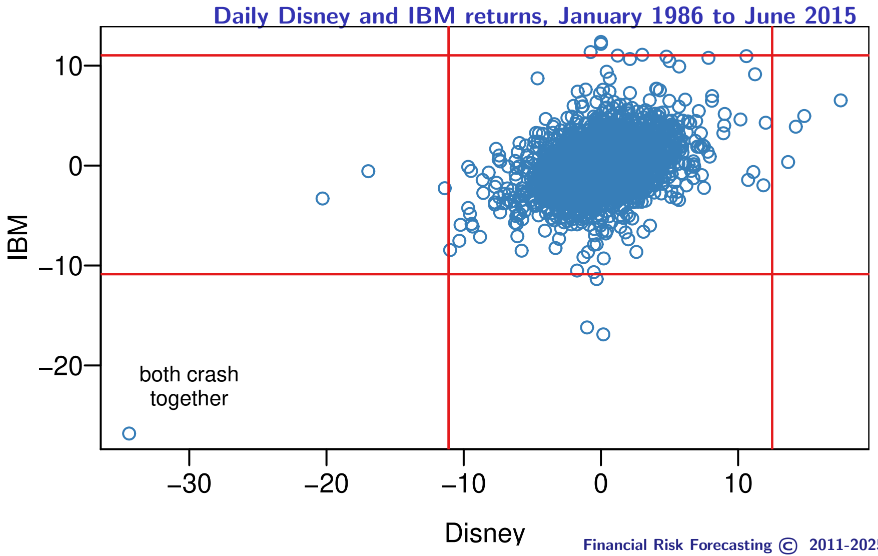
Application of Copulas

- To illustrate we use the same data on Disney and IBM as used before
- By comparing a scatterplot for simulated bivariate normal data with one for the empirical data, we see that the two do not have the same joint extremes

Gaussian Scatterplot



Empirical Scatterplot

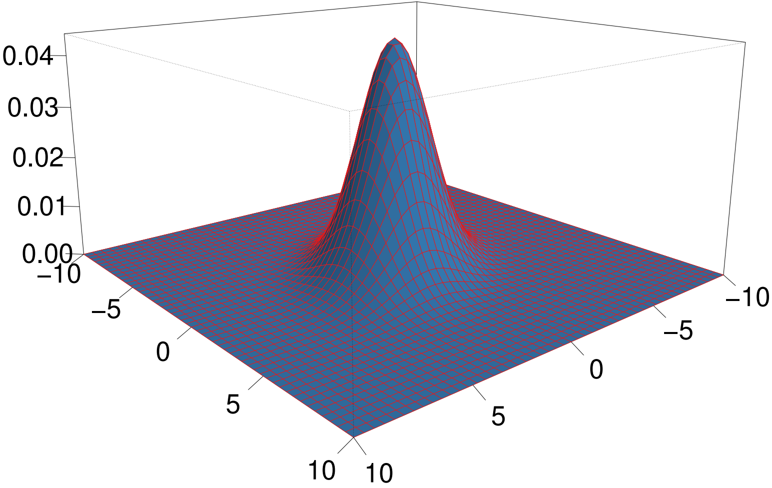


Application of Copulas

- We estimate two copulas for the data, a Gaussian copula and a Student-t copula
- The copulas can be drawn in three dimensions

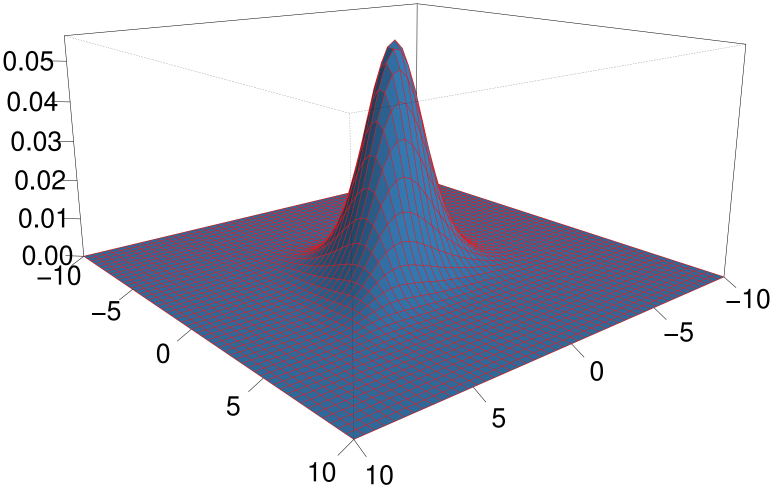
Fitted Gaussian Copula

Daily Disney and IBM returns, January 1986 to June 2015



Fitted Student-t Copula

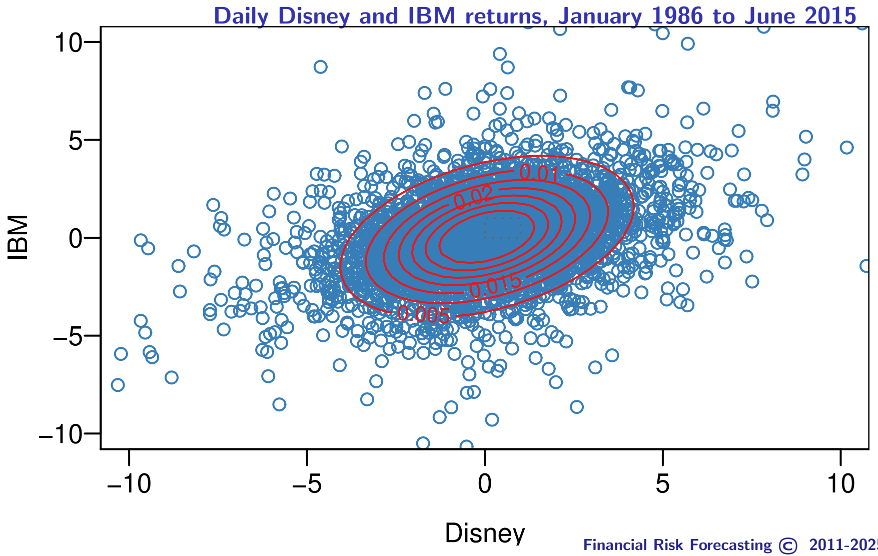
Daily Disney and IBM returns, January 1986 to June 2015



Application of Copulas

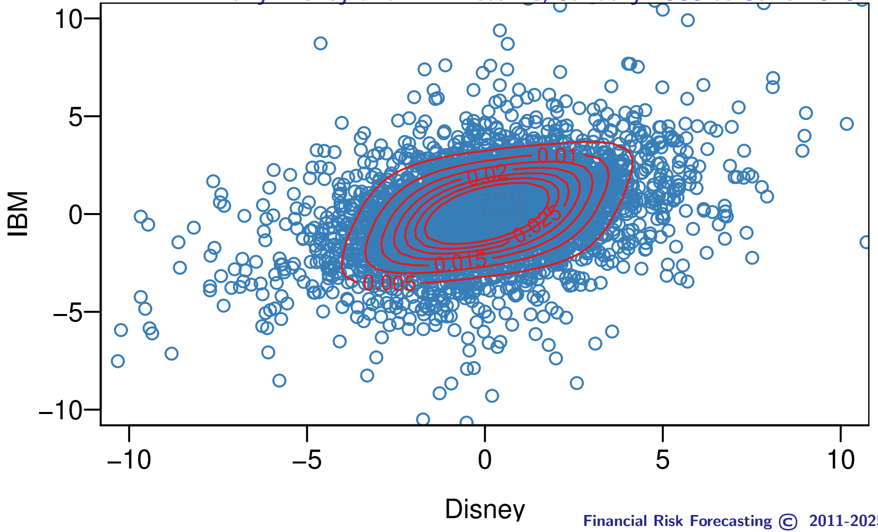
- It can be difficult to compare distributions by looking at three-dimensional graphs
- Contour plots may give a better comparison

Contours of Gaussian Copula



Contours of Student-t Copula

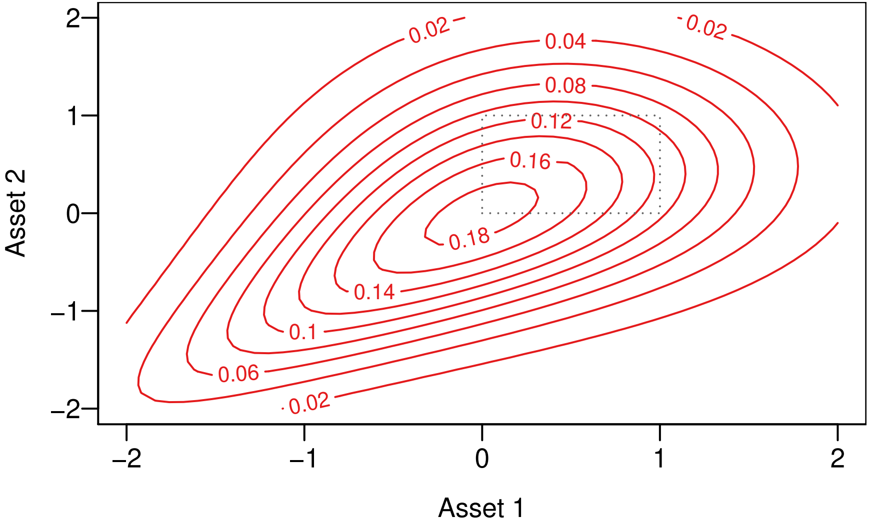
Daily Disney and IBM returns, January 1986 to June 2015



Clayton's Copula

- As noted earlier, there are a number of copulas available
- One widely used is the Clayton copula, which allows for asymmetric dependence
- Parameter θ measures the strength of dependence
- We estimate a Clayton copula for the same data as before

Contours of Clayton's Copula, $\theta = 1$



Contours of Clayton's Copula, $\theta = 0.483$